

Estimation and inference in nonlinear nonstationary panel data models

Citation for published version (APA):

Wan, L. (2012). *Estimation and inference in nonlinear nonstationary panel data models*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20120209lw>

Document status and date:

Published: 01/01/2012

DOI:

[10.26481/dis.20120209lw](https://doi.org/10.26481/dis.20120209lw)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Estimation and Inference
in
Nonlinear Nonstationary Panel Data Models

Lei Wan

© Lei Wan, 2012

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

This book was typeset by the author using L^AT_EX.

Published by Universitaire Pers Maastricht

ISBN: 978-94-6159-119-7

Printed in The Netherlands by Datawyse Maastricht

Estimation and Inference
in
Nonlinear Nonstationary Panel Data Models

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
donderdag 9 februari 2012 om 16.00 uur

door

Lei Wan



Promotores:

Prof. dr. F.C. Palm

Prof. dr. J.R.Y.J Urbain

Beoordelingscommissie:

Prof. dr. A. Monfort (voorzitter)

Prof. dr. J. Gonzalo (Universidad Carlos III de Madrid)

Prof. dr. P. Schotman

Dit onderzoek werd gefinancierd door de Maastricht Research School of Economics of Technology and Organizations (METEOR).

Acknowledgements

Looking back, I have never regretted the choice of pursuing a Ph.D. degree. The experience is exciting and enjoyable. Honestly, there have been the moments of frustration when I got stuck, or things just did not seem to work as they were expected. However, statistically, the number of these frustrating moments is highly nonsignificant if stationarity has been assumed. I have to say that, after hard working, I have truly enjoyed every single moment no matter there are interesting results being obtained or new problems being discovered.

Many people have contributed significantly to my enjoyment over the recent years, and I would like to take the opportunity here to thank some of them. First of all, I am very grateful to my supervisors Franz Palm and Jean-Pierre Urbain. They are not only knowledgeable but also considerate. They set themselves a good example to me on taking right attitude and being responsible as researchers, which, I believe, not every Ph.D. student would have the chance to experience. It has truly been a privilege for me to work under their supervision. I could not think about any better supervision. Furthermore, I would like to thank Jean-Pierre for his teaching which attracts me deeply toward theoretical econometrics.

I would like to thank the members of the reading committee, Alain Monfort, Jesús Gonzalo and Peter Schotman, for their time and careful reading of the manuscript. I thank all my colleagues at the Department of Quantitative Economics for providing the pleasant and inspiring working atmosphere, and the decent education to me when I was still a student. I am also grateful to all colleagues sitting at the 4th floor of the main building and Room 2002 of TS 49, who provided motivating discussion and comments. In particular, I would like to thank Stephan Smeekes, who personally contributed to this thesis in form of the Dutch summary.

Furthermore, I would like to thank my family and friends. My parents have always supported and encouraged me in all possible ways. I would like to thank my girlfriend Changjin Feng who has

never made a single complain about my Ph.D. working nights and weekends.

Last but not least, I would like to thank all the authors listed in the Bibliography, who have accompanied me day and night on producing this thesis.

Lei Wan

Maastricht, 2011

Contents

Acknowledgements	v
1 Introduction	1
1.1 Objective of the Thesis	1
1.2 Motivation	2
1.3 Contributions and Thesis Structure	3
2 Nonlinear Transformations of Integrated Time Series, Co-summability, Non-linear Asymptotics for Stationary Processes	7
2.1 Introduction	7
2.2 Nonlinear Transformations of Integrated Time Series	10
2.2.1 Functions of Integrated Processes & Local Time of Brownian Motions	10
2.2.2 Asymptotics for Sample Moments	14
2.2.3 Extensions	17
2.3 Co-summability: Extension of Linear Cointegration	19
2.4 Nonlinear Asymptotic Theory for Stationary Processes	23
2.4.1 Consistency	23
2.4.2 Asymptotic Distribution	25
2.5 Conclusion	28
3 Pooled Nonlinear Least Squares Estimation for Independent Co-summable Panels	33
3.1 Introduction	33
3.2 The Model and Assumptions	35
3.3 Simple Strong Co-summable Regressions	41
3.3.1 Preliminary Results	42
3.3.2 Consistency	43
3.3.3 Asymptotic Normality	46

3.3.4	Estimation of the Covariance Matrix	49
3.4	Multiple Strong Co-summable Regressions	50
3.4.1	Consistency	52
3.4.2	Asymptotic Normality	55
3.4.3	Estimation of the Covariance Matrix	58
3.5	Smooth Transition Co-summable Regressions	59
3.5.1	Consistency and Asymptotic Normality	60
3.5.2	Estimation of the Covariance Matrix	62
3.6	Monte Carlo Simulations	63
3.7	Conclusion	67
3.A	Tables	70
3.B	Figures	75
3.C	Appendix	83
3.C.1	Lemma 3.C.1	83
3.C.2	Proof of Lemma 3.C.1	83
3.C.3	Lemma 3.C.2	90
3.C.4	Proof of Lemma 3.C.2	90
3.C.5	Proof of Lemma 3.1	91
3.C.6	Proof of Lemma 3.2	92
3.C.7	Proof of Lemma 3.3	92
3.C.8	Proof of Lemma 3.4	93
3.C.9	Proof of Theorem 3.1	95
3.C.10	Proof of Theorem 3.2	96
3.C.11	Proof of Lemma 3.5	97
3.C.12	Proof of Theorem 3.3	97
3.C.13	Proof of Theorem 3.4	101
3.C.14	Proof of Corollary 3.1	103
3.C.15	Proof of Corollary 3.2	103
3.C.16	Proof of Theorem 3.5	103
3.C.17	Proof of Theorem 3.6	106
3.C.18	Proof of Theorem 3.7	111
3.C.19	Proof of Theorem 3.8	115

3.C.20 Proof of Corollary 3.3	118
3.C.21 Proof of Corollary 3.4	119
3.C.22 Proof of Corollary 3.5	119
3.C.23 Proof of Corollary 3.6	121
3.C.24 Proof of Corollary 3.7	121
3.C.25 Proof of Corollary 3.8	121
4 Nonlinear Least Squares Dummy Variable Estimation for Independent Co-	
summable Panels with Fixed Effect	123
4.1 Introduction	123
4.2 The Model and Assumptions	124
4.3 Simple Strong Co-summable Regressions	130
4.3.1 Consistency	130
4.3.2 Asymptotic Normality	131
4.3.3 Estimation of the Covariance Matrix	132
4.4 Multiple Strong Co-summable Regressions	133
4.4.1 Consistency	134
4.4.2 Asymptotic Normality	136
4.4.3 Estimation of the Covariance Matrix	138
4.5 Smooth Transition Co-summable Regressions	138
4.5.1 Consistency and Asymptotic Normality	139
4.5.2 Estimation of the Covariance Matrix	141
4.6 Monte Carlo Simulations	142
4.7 Conclusions	144
4.A Tables	146
4.B Figures	151
4.C Appendix	157
4.C.1 Proof of Theorem 4.1	157
4.C.2 Proof of Theorem 4.2	158
4.C.3 Proof of Theorem 4.3	159
4.C.4 Proof of Theorem 4.4	161
4.C.5 Proof of Corollary 4.1	162
4.C.6 Proof of Corollary 4.2	162

4.C.7 Proof of Theorem 4.5	162
4.C.8 Proof of Theorem 4.6	163
4.C.9 Proof of Theorem 4.7	163
4.C.10 Proof of Theorem 4.8	163
4.C.11 Proof of Corollary 4.3	163
4.C.12 Proof of Corollary 4.4	163
4.C.13 Proof of Corollary 4.5	164
4.C.14 Proof of Corollary 4.6	164
4.C.15 Proof of Corollary 4.7	164
4.C.16 Proof of Corollary 4.8	164
5 Polynomial Co-summable Panels with Common Factors	165
5.1 Introduction	165
5.2 The Model and Assumptions	166
5.3 The CCEP Estimator and Asymptotic Results	171
5.4 Monte Carlo Simulations	173
5.5 Conclusion	175
5.A Tables	177
5.B Appendix	180
5.B.1 Proof of Proposition 5.1: Heuristic Arguments	180
5.B.2 Proof of Proposition 5.2: Heuristic Arguments	181
6 Conclusion and Further Research	183
6.1 Summary, Limitations and Further Research	183
Bibliography	187
Nederlandse samenvatting	193
Curriculum Vitae	197

List of Tables

2.1	$I - \text{regular}$ Functions	12
2.2	$H - \text{regular}$ Functions	13
3.1	MAE for Pooled NLS: $\hat{\alpha}_{n,T}$	70
3.2	MAE for Pooled NLS: $\hat{\beta}_{n,T}$	70
3.3	MAE for Pooled NLS: $\hat{\gamma}_{n,T}$	70
3.4	MAE for Pooled NLS: $\hat{c}_{n,T}$	71
3.5	Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{1,n,T}$	71
3.6	Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{2,n,T}$	71
3.7	Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{3,n,T}$	72
3.8	Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{4,n,T}$	72
3.9	Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{5,n,T}$	72
3.10	MAE for Pooled NLS: $\hat{\alpha}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	73
3.11	MAE for Pooled NLS: $\hat{\beta}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	73
3.12	MAE for Pooled NLS: $\hat{\gamma}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	73
3.13	MAE for Pooled NLS: $\hat{c}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	74
4.1	MAE for Pooled NLSDV: $\hat{\alpha}_{n,T}$	146
4.2	MAE for Pooled NLSDV: $\hat{\beta}_{n,T}$	146
4.3	MAE for Pooled NLSDV: $\hat{\gamma}_{n,T}$	146
4.4	MAE for Pooled NLSDV: $\hat{c}_{n,T}$	147
4.5	Estimated Variance for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{1,n,T}$	147
4.6	Estimated Variance for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{2,n,T}$	147
4.7	Estimated Variance for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{3,n,T}$	148
4.8	Estimated Variance for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{4,n,T}$	148
4.9	Estimated Variance for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{5,n,T}$	148
4.10	MAE for Pooled NLSDV: $\hat{\alpha}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	149

4.11	MAE for Pooled NLSDV: $\hat{\beta}_{n,T}, COV(x_{i,t}, z_{i,t}) = 0$	149
4.12	MAE for Pooled NLSDV: $\hat{\gamma}_{n,T}, COV(x_{i,t}, z_{i,t}) = 0$	149
4.13	MAE for Pooled NLSDV: $\hat{c}_{n,T}, COV(x_{i,t}, z_{i,t}) = 0$	150
5.1	Heterogeneous Coefficients: CCEP $\beta_2, I(1) v_{i,t}$	177
5.2	Heterogeneous Coefficients: CCEP $\beta_1, I(1) v_{i,t}$	177
5.3	Heterogeneous Coefficients: CCEP $\beta_2, I(0) v_{i,t}$	177
5.4	Heterogeneous Coefficients: CCEP $\beta_1, I(0) v_{i,t}$	178
5.5	Homogeneous Coefficients: CCEP $\beta_2, I(1) v_{i,t}$	178
5.6	Homogeneous Coefficients: CCEP $\beta_1, I(1) v_{i,t}$	178
5.7	Homogeneous Coefficients: CCEP $\beta_2, I(0) v_{i,t}$	179
5.8	Homogeneous Coefficients: CCEP $\beta_1, I(0) v_{i,t}$	179

List of Figures

3.1	Estimated PDF for Pooled NLS: $n = 10$, α , $\hat{\Sigma}_{1,n,T}$	75
3.2	Estimated PDF for Pooled NLS: $n = 10$, α , $\hat{\Sigma}_{2,n,T}$	75
3.3	Estimated PDF for Pooled NLS: $n = 10$, β , $\hat{\Sigma}_{1,n,T}$	76
3.4	Estimated PDF for Pooled NLS: $n = 10$, β , $\hat{\Sigma}_{2,n,T}$	76
3.5	Estimated PDF for Pooled NLS: $n = 10$, γ , $\hat{\Sigma}_{1,n,T}$	77
3.6	Estimated PDF for Pooled NLS: $n = 10$, γ , $\hat{\Sigma}_{2,n,T}$	77
3.7	Estimated PDF for Pooled NLS: $n = 10$, c , $\hat{\Sigma}_{1,n,T}$	78
3.8	Estimated PDF for Pooled NLS: $n = 10$, c , $\hat{\Sigma}_{2,n,T}$	78
3.9	Realizations: $n = 10$, $T = 30$, $COV(x_{i,t}, z_{i,t}) > 0$	79
3.10	Realizations: $n = 10$, $T = 30$, $COV(x_{i,t}, z_{i,t}) < 0$	79
3.11	Realizations: $n = 10$, $T = 30$, $COV(x_{i,t}, z_{i,t}) = 0$	80
3.12	Estimated PDF for Pooled NLS: $n = 10$, α , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	80
3.13	Estimated PDF for Pooled NLS: $n = 10$, β , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	81
3.14	Estimated PDF for Pooled NLS: $n = 10$, γ , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	81
3.15	Estimated PDF for Pooled NLS: $n = 10$, c , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	82
4.1	Estimated PDF for Pooled NLSDV: $n = 10$, α , $\hat{\Sigma}_{1,n,T}$	151
4.2	Estimated PDF for Pooled NLSDV: $n = 10$, α , $\hat{\Sigma}_{2,n,T}$	151
4.3	Estimated PDF for Pooled NLSDV: $n = 10$, β , $\hat{\Sigma}_{1,n,T}$	152
4.4	Estimated PDF for Pooled NLSDV: $n = 10$, β , $\hat{\Sigma}_{2,n,T}$	152
4.5	Estimated PDF for Pooled NLSDV: $n = 10$, γ , $\hat{\Sigma}_{1,n,T}$	153
4.6	Estimated PDF for Pooled NLSDV: $n = 10$, γ , $\hat{\Sigma}_{2,n,T}$	153
4.7	Estimated PDF for Pooled NLSDV: $n = 10$, c , $\hat{\Sigma}_{1,n,T}$	154
4.8	Estimated PDF for Pooled NLSDV: $n = 10$, c , $\hat{\Sigma}_{2,n,T}$	154
4.9	Estimated PDF for Pooled NLSDV: $n = 10$, α , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	155
4.10	Estimated PDF for Pooled NLSDV: $n = 10$, β , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$	155

4.11	Estimated PDF for Pooled NLSDV: $n = 10, \gamma, \hat{\Sigma}_{1,n,T}, COV(x_{i,t}, z_{i,t}) = 0$ 156
4.12	Estimated PDF for Pooled NLSDV: $n = 10, c, \hat{\Sigma}_{1,n,T}, COV(x_{i,t}, z_{i,t}) = 0$ 156

Introduction

1.1 Objective of the Thesis

The analysis of nonstationary time series has already been one of the major research topics in time series econometrics for decades. Empirical studies indicate that data on variables such as real GDP, inflation, exchange rates and stock price, in general, show evidence of being $I(1)$ or integrated¹. The long run equilibrium relationship among such integrated variables is defined as cointegration by Engle and Granger (1987). In 2003, Clive Granger and Robert Engle were awarded The Nobel Prize in Economics for their contribution on the analysis of nonstationary time series. Recently, there have been many empirical studies on linear panel data models with integrated time series components, which have been facilitated by the construction and availability of many important panel data sets covering different individuals or countries. Many of these panel data sets possess both a large cross-section data (n) dimension and a large number of time series (T) observations. Such cases have motivated the recent advance in the theoretical work on large n and large T asymptotics.

This thesis focuses on panel data models with nonlinear components of integrated time series processes. The objective is to derive and analyze the large n and large T asymptotics for some nonlinear estimators, including the *Pooled Nonlinear Least Squares* estimator, the *Pooled Nonlinear Least Squares Dummy Variable* estimator and the modified *Common Correlated Effects Pooled* estimator, based on pooling the data. In particular, we combine the newest concepts and latest techniques introduced in the literature on nonlinear nonstationary time series analysis to motivate and interpret panel co-summable regression models, and analyze related estimation techniques for the purpose of inference. We mainly consider nonlinear models for cross-sectionally independent nonstationary panels with strictly exogenous independent variables, which cover a number of empirically popular

¹The type of nonstationarity specified in this thesis is virtually $I(1)$ or integrated. Hence, later in the thesis, the terms- $I(1)$, integrated and nonstationary-are interchangeably used without any confusion. Furthermore, the type of stationarity considered in this thesis is based on the definition of $I(0)$ processes introduced by Davidson (2009).

nonlinear models such as Logistic Smooth Transition models and Polynomial Cointegration models². Our results show that the above mentioned pooled nonlinear estimators converge to a *Normal* distribution, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$, but the rates of convergence are \sqrt{n} multiplied by a factor of T^d , where $d > 1/2$ depending on the type of nonlinearity. Moreover, we adopt a common factor structure for Panel Polynomial Co-summable Regression models to analyze cross-sectionally dependent panels. It is worth noticing that the results obtained in this thesis are about nonlinear transformations of $I(1)$ processes, which should be distinguished from the branch of literature considering nonlinear transformations of linear cointegration equations, for example, the type of nonlinear error correction models in Escribano and Mira (2001) and Bec and Rahbek (2004), where they consider the nonlinear transformations of $I(0)$ processes.

1.2 Motivation

In the empirical econometric literature, it has been a long history to treat macroeconomic variables as being integrated of order one $I(1)$. If in the long run these variables move in line with each other, they are assumed to be cointegrated. Theoretically, the statistical analysis of models with $I(1)$ components requires different techniques from those conventionally used by the classical econometric analysis and the stationary time series analysis. Both practical applications and theoretical analysis of models with $I(1)$ variables turn out to be very important topics in the recent econometric literature, for example, to study the behavior of real per capita output growth, to test hypotheses such as the economic hypothesis of purchasing power parity, to test for cointegration and so on. A fairly complete asymptotic theory for linear regressions with $I(1)$ components has been developed as the basis of ongoing empirical work, especially on the subject of unit root testing and cointegration modeling.

On the other hand, economic theory seems to suggest that the relationships between many macroeconomic variables are very likely nonlinear. Intuitively, the nonlinearity might be due to economic cycles or changes of policy regimes. A number of empirical applications of nonlinear regression models, including the Threshold model, the Smooth Transition model and the Markov Switching model, can be found in the literature of international economics, such as Balke and Wohar (1998), Baum, Barkoulas and Caglayan (2001), Enders and Falk (1998), Lo and Zivot (2001), O'Connell (1998), Taylor (2001) and González, Teräsvirta and Van Dijk (2005). However, the analysis of nonlinear

²Later in this thesis, we reinterpret this type of nonlinear models as the *Panel Polynomial Co-summable Regression* model based on concepts introduced by Rico (2009) and Rico and Gonzalo (2011).

regression models based merely on time series data sets seems not efficient enough to identify the nonlinearity in many occasions. For example, in the case of Smooth Transition models, the time series data sets generally contain few observations on the transition stage in comparison to the number of observations on different economic regimes. In this case, by pooling the observations in panel data sets, more information can be obtained for parameters of interest if the sets of parameters are assumed to have common values across units. Hence, the type of models and estimation methods based on pooling the observations together seem appealing to efficiently identify and model nonlinear economic relationships.

In the literature on linear dynamic panels, fruitful theoretical work has been done, such as Phillips and Moon (1999b) and Pedroni (2001), on the large n and large T asymptotics. In this thesis, we try to combine the elements of nonstationarity, nonlinearity, panel data and pooled estimation techniques, and study large n and large T asymptotics for nonlinear panel models with nonstationary regressors. In order to interpret the type of the nonstationarity generated by the nonlinear transformations of $I(1)$ variables, we adopt the concepts of *summability* and *co-summability*, which are recently introduced by Rico (2009) and Rico and Gonzalo (2011), to extend the concepts of integration and cointegration from linear situations to nonlinear situations. The underlying theoretical work is mainly based on the recent contribution in studying multivariate nonlinear regression models of nonstationary time series by Park and Phillips (1999), Park and Phillips (2001), Chang, Park and Phillips (2001), De Jong (2002) and Saikkonen and Choi (2004), as well as those in linear panel regression models by Phillips and Moon (1999b), Pesaran (2006) and Kapetanios, Pesaran and Yamagata (2011). Given the fact that each chapter of this thesis contains a discussion of the relevant literature, no review is provided here.

1.3 Contributions and Thesis Structure

The contribution of this thesis lies in first attempt to extending linear panel estimation and inference methods to nonlinear panel models and deriving the large n and large T asymptotics. The main contribution is based on cross-sectionally independent panels. Finally, it is extended to a special type of nonlinear panels with common factors. In Chapter 2 we provide a detailed overview of the theoretical tools required for the analysis of nonlinear nonstationary panel data models with large n and large T dimensions. The tools include existing asymptotic results for nonlinear transformations of a $I(1)$ process, summability and co-summability, and the nonlinear asymptotic theory for stationary processes. The study about nonlinear transformations of nonstationary time series has been initiated

by Granger and Hallman (1988), which is refined and published as Granger and Hallman (1991). Later on, the nonlinearly transformed processes are extensively studied as long memory processes by Ermini and Granger (1993), Corradi (1995) and Franses and McAleer (1996). Very recently, Rico (2009) and Rico and Gonzalo (2011) introduce the concept of summability based on the asymptotic order of the partial sum of a time series process. Accordingly, the long run relationship between processes with the same order of summability is defined as co-summability. Two out of the three classes of nonlinear functions categorized in Park and Phillips (1999), Park and Phillips (2001), Chang et al. (2001) on nonlinear regression models with integrated regressors are analyzed, namely *integrable functions* and *asymptotically homogeneous functions*. The local time of a Brownian motion turns out to be an important concept on which the asymptotic property of integrable functions heavily relies. De Jong (2002), Saikkonen and Choi (2004), Kasparis (2008), Hong and Wagner (2008) and Hong and Wagner (2011) discuss estimation and inference issues under the presence of endogenous regressors. In the stationary case, the Law of Large Numbers and Central Limit Theory have been shown for various nonlinear transformations if the appropriate assumptions are imposed, i.e., Pötscher and Prucha (1997). Finally, potential open issues are summarized as points for future research.

In Chapter 3, we first derive the asymptotic properties of the time series sample moments defined by the product of integrable functions and asymptotically homogeneous functions. Then, we define separately the concepts of *simple strong co-summability* and *multiple strong co-summability*. Accordingly, we introduce, based on integrable functions and asymptotic homogeneous functions, different types of simple strong co-summable regression models and multiple strong co-summable regression models for cross-sectionally independent panels with both large n and large T dimensions. Furthermore, the existence of the Pooled Nonlinear Least Squares estimator is investigated under the assumption of compactness on the parameter space. As $T \rightarrow \infty$ followed by $n \rightarrow \infty$, We also prove the consistency of the Pooled Nonlinear Least Squares estimator and derive its asymptotic distribution. Several consistent estimators of the covariance matrix are proposed for inference purposes. Finally, we take the panel Logistic Smooth Transition Co-summable Regression model as an example to obtain its asymptotics and investigate the finite sample performance with Monte Carlo simulations.

Chapter 4 augments the nonlinear panel co-summable models studied in Chapter 3 with an additive fixed effect term for each cross-sectional individual. For such models, the estimator used by González et al. (2005) is formalized as the Pooled Nonlinear Least Squares Dummy Variable estimator. The estimator is proved to be consistent and its asymptotic distribution is derived for

simple strong co-summable regressions and multiple strong co-summable regressions respectively. The Pooled Nonlinear Least Squares Dummy Variable estimator requires a set of identification conditions different from that required by the Pooled Nonlinear Least Squares estimator. Further results are obtained for the panel Logistic Smooth Transition Co-summable Regression model, and consistent estimators for the covariance matrix are proposed.

In Chapter 5, we consider a special case of panel multiple strong co-summable regression models, the panel polynomial co-summable regression models, with additive unobserved $I(1)$ common factors. Moreover, we modify the Common Correlated Effects Pooled estimator introduced by Pesaran (2006) to estimate the common parameters or the common parts of the heterogeneous parameters. We consider both the case where the regressors are linearly cointegrated with the factors and the case where there is no presence of such linear cointegration relationships. We use Monte Carlo simulations and a heuristic but incomplete theoretical analysis to investigate the finite sample performance and the asymptotic properties of the modified Common Correlated Effects Pooled estimator for the coefficients of interest being assumed to be homogeneous or heterogeneous across individual units.

Chapter 6 summarizes the main findings of the thesis, discusses the limit of the current work and presents some outlines for further research.

Nonlinear Transformations of Integrated Time Series, Co-summability, Nonlinear Asymptotics for Stationary Processes

2.1 Introduction

This chapter provides an overview of the theoretical tools available in the literature and required for the analysis of nonlinear nonstationary panel data models, which include asymptotic theory for nonlinear nonstationary time series models, summability and co-summability, and the asymptotic theory of \mathbf{M} -estimators¹ for dynamic nonlinear regression models with $I(0)$ regressors. Given that the first two topics are relatively young in the econometric literature, we expect that this chapter provides necessary preliminaries that could prepare readers for a better understanding of the later chapters². In addition, we intend to point out and summarize potential open issues for further research, and motivate possible modification and integration of the existing results for applications such as the panel Smooth Transition Regression (STR) models.

According to the author's knowledge, the discussion on nonlinear transformations of nonstationary time series was initiated by Granger and Hallman (1988), in the econometric literature, which was later on refined and published as Granger and Hallman (1991). Granger and Hallman (1991) studied the time series properties of the first order difference of some nonlinear transformations, $f(x_t)$ with x_t being an $I(1)$ process, by applying a mean value expansion to some given nonlinear transformations. They showed that the process of the first order difference of $f(x_t)$, including polynomial functions,

¹According to Pötscher and Prucha (1997), \mathbf{M} -estimators are defined as the optimal solutions of an objective function and include Nonlinear Least Squares estimators, Nonlinear Maximum Likelihood estimators, Generalized Method of Moments estimators and so on.

²Some of the remarks may contain subtle refinements and further discussion on issues in the most recent literature. These are included for completeness and readers being interested. Hence, the remarks can be skipped without hindering any understandings on later chapters.

indicator functions, trigonometric functions, exponential functions, logarithm functions and reciprocal functions, fail the set of sufficient conditions for the Functional Central Limit Theorem (FCLT). Hence, $f(x_t)$ is not $I(1)$. Their simulation studies showed that the Dickey-Fuller (DF) test and the Augmented Dickey-Fuller (ADF) test are only powerful to detect this property for some functions such as indicator functions, exponential functions and trigonometric functions but not others, i.e., polynomial functions. On the other hand, their simulation studies on the autocorrelation functions of the levels of $f(x_t)$ showed that most of the nonlinearly transformed processes have autocorrelation structures similar to the $I(1)$ process. All in all, it seems that the conventional definitions and classification of $I(1)$ and $I(0)$ processes in time series analysis are not sufficient to describe the properties of $f(x_t)$. Therefore, Granger and Hallman (1988) introduced the definitions of long memory processes and short memory processes, which is basically an extension of the linear measurement concept-autocorrelation-to a more general temporal, possibly nonlinear, dependence. Moreover, Granger and Hallman (1988) and Granger and Hallman (1991) proposed the Rank Dickey-Fuller test and the Augmented Rank Dickey-Fuller test for the null hypothesis that assumes the underlying process to be a strictly monotone transformation of an $I(1)$ process. The finite sample performance of these tests was studied by simulations but the asymptotic properties were not developed. They also showed that x_t and $f(x_t)$ cannot be cointegrated unless the function $f(\cdot)$ is affine. Finally, they conclude that if x_t and z_t are cointegrated, $f(x_t)$ and $f(z_t)$ can also be cointegrated if either $f(\cdot)$ is homogeneous or the cointegrating coefficient for x_t and z_t is 1.

Ermini and Granger (1993) adopted the Hermite polynomial approach and obtained analytical results for the autocorrelation structure respectively of the polynomial transformation, the exponential transformation, the periodic transformation and the neural-network transformation of an $I(1)$ processes. In general, the polynomial transformation exhibits a time trend and the autocorrelation structure preserves the long memory property of $I(1)$ processes. In small sample, the high-degree polynomial transformation may appear to be $I(0)$. The exponential transformation exhibits $I(0)$ autocorrelation structure and may appear to be deterministic depending on the drift in x_t . The periodic transformation behaves as stationary, zero-mean, homoscedastic $AR(1)$ processes, for large samples, even if the drift in x_t is not zero. The neural-network transformation basically has an autocorrelation structure similar to that of an $I(1)$ process. Conclusively, Granger and Hallman (1988), Granger and Hallman (1991) and Ermini and Granger (1993) showed that the autocorrelation functions is not sufficient and reliable to distinguish long memory processes from short memory processes. Hence, Corradi (1995) adopted the definition of strong mixing to characterize short memory. Besides, Cor-

radi (1995) exploited the martingale property of $I(1)$ processes by applying the Doob decomposition. The monotonicity and convexity of the nonlinear transformation are critical for the results. In particular, strict convex (concave) transformations display the properties of an $I(1)$ component, but the first order difference of such transformations are not necessarily strong mixing (short memory). Based on the conclusion of Corradi (1995), Franses and McAleer (1996) showed that the standard DF test or ADF test are not valid and powerful enough for detecting the $I(1)$ component in the logarithm transformation of an $I(1)$ process. They proposed a modified ADF test based on the Taylor expansion of the nonlinear function and showed that the test is robust against Box-Cox type nonlinear transformations.

On the one hand, the set of conditions satisfied by the short memory processes seem to be limited to the set of sufficient conditions for the Law of Large Numbers (LLNs) and Central Limit Theory (CLT), such as martingale difference properties, mixing properties or the Near Epoch Dependence (NED) property. On the other hand, most of the above mentioned literature seems to show that the set of sufficient conditions for the FCLT are not sufficient to describe the long memory processes. However, considering the nonlinear transformations of $I(1)$ processes, the definitions of short memory and long memory in the literature seem not exhaustive, in particular, the ones studied by Corradi (1995). Hence, not only asymptotic results other than the LLNs, CLT and FCLT are expected, but also more appropriate standards or definitions for classifying different nonlinear transformations $f(x_t)$ are required. The major parts of this chapter are devoted to surveying the new developments in the literature concerning the previously mentioned two needs.

The chapter is organized as follows. Section 2.2 summarizes existing asymptotic results for nonlinear regression models with integrated time series regressors in the literature. Section 2.3 introduces the definitions of summability and co-summability, which extend the definitions of integration and cointegration from the linear case to the nonlinear case. Just as how we define the balancedness of a linear regression, these concepts are necessary for a correct nonlinear specification. Section 2.4 reviews the classical asymptotic theory of \mathbf{M} -estimators for nonlinear regression models with stationary regressors. Attention is paid to the derivation of the consistency and the asymptotic distribution. Some concluding remarks are made in Section 2.5. In particular, potential open issues are summarized.

A word about the notations. For a vector $x = (x_i)$ or a matrix $A = (a_{ij})$, the modulus $|\cdot|$ is taken element by element. Therefore, $|x| = (|x_i|)$ and $|A| = (|a_{ij}|)$. The standard Euclidean norm of a vector is denoted by $\|\cdot\|$, i.e. $\|x\|^2 = \sum_i (x_i)^2$. For a matrix, $\|A\|$ signifies the operator norm defined

by $\|A\| = \sup_x \|Ax\|/\|x\|$. The maximum of the moduli is denoted by $\overline{\|\cdot\|}$, i.e., $\overline{\|x\|} = \max_i |x_i|$ and $\overline{\|A\|} = \max_{i,j} |a_{ij}|$. Moreover, the k by k dimension identity matrix is denoted by I_k . For a function, which can be vector- or matrix-valued, $\overline{\|\cdot\|}_K$ signifies the supremum over a subset K of its domain, so that $\overline{\|f\|}_K = \sup_{x \in K} \|f(x)\|$. The subset K , over which the supremum is taken, will not be specified if it is clear from the context. Furthermore, the indicator function is written as $1\{\cdot\}$. Letters such as p, q, m, l and δ generally denote positive integers and the definitions should be clear under the context. \equiv_d denotes equality in distribution. In addition, standard terminologies and notations in probability and measure theory are used throughout the rest of the chapter such as almost sure convergence ($\xrightarrow{a.s.}$), convergence in probability (\xrightarrow{p}), convergence in distribution (\xrightarrow{d}) and weak convergence (\Rightarrow). $W(r)$ denotes a vector of standard Brownian motions and $B(r)$ denotes a vector of general Brownian motions. Their dimensions are clear from the context. For a function $f(x, \theta)$, \dot{f} , \ddot{f} and \dddot{f} denote, respectively, the first, second and third derivatives of f with respect to θ . They are possibly vectors and arranged by the lexicographic ordering of their indices based on the vector θ .

2.2 Nonlinear Transformations of Integrated Time Series

This section is based on the recent developments and extensions initiated by Park and Phillips (1999), Park and Phillips (2001) and Chang et al. (2001) on nonlinear regression models where the regressors are assumed to be integrated time series processes. Two out of the three classes of nonlinear functions categorized in Park and Phillips (2001) are introduced, namely integrable functions and asymptotically homogeneous functions which are respectively defined as I -regular and H -regular³. Moreover, the local time of Brownian motion tends to be an important concept on which the asymptotic properties of I -regular functions heavily rely.

2.2.1 Functions of Integrated Processes & Local Time of Brownian Motions

In the stationary case, the LLNs and CLT have been shown to be still valid for various nonlinear transformations if the appropriate assumptions are imposed, i.e., Pötscher and Prucha (1997). However, in the nonstationary case, the asymptotic behavior of nonlinear functions of integrated processes is proved to be very different and more dependent on the nonlinear properties of the chosen functions. Hence, the following two main categories of nonlinear functions are considered in the

³The H -regular function considered here and later on is actually the H_0 -regular function defined by Park and Phillips (2001)

literature: I – regular functions and H – regular functions. H – regular functions are based on the family of regular functions which are defined by Park and Phillips (2001) as follows:

Definition 2.1. A family of $f(\cdot, \theta)$ is said to be regular on Θ if and only if,

- (a) for all $\theta \in \Theta$, $f(\cdot, \theta)$ is continuous in a neighborhood of infinity, and
- (b) for all $\theta \in \Theta$ and any compact set K of \mathbb{R} given, there exist for each $\epsilon > 0$ continuous functions $\underline{f}_\epsilon, \bar{f}_\epsilon$ and $\delta_\epsilon > 0$ such that $\underline{f}_\epsilon(x, \theta) \leq f(y, \theta) \leq \bar{f}_\epsilon(x, \theta)$ for all $|x - y| < \delta_\epsilon$ on K , and such that $\int_K (\bar{f}_\epsilon(x, \theta) - \underline{f}_\epsilon(x, \theta)) dx \rightarrow 0$ as $\epsilon \rightarrow 0$, and
- (c) for all $x \in \mathbb{R}$, $f(x, \cdot)$ is equicontinuous in a neighborhood of x .

Remark 2.1.

- (1) Any regular function is locally integrable which guarantees the existence of the limiting distribution. A vector- or matrix-valued function is called regular when each of its components is regular. Park and Phillips (1999) pointed out that locally bounded monotone functions and continuous functions are regular. Logarithm functions and reciprocal functions are not regular.
- (2) In addition, Lemma A1 of Park and Phillips (2001) shows that the space of regular functions is closed under the usual operations of addition, subtraction and multiplication. This class of functions is used for deriving the asymptotics for transformations of normalized $I(1)$ processes⁴.

Definition 2.2. A family of functions $f(\cdot, \theta)$ is said to be I – regular on Θ if,

- (a) for each $\theta_0 \in \Theta$, there exists a neighborhood N_0 of θ_0 and $g : \mathbb{R} \rightarrow \mathbb{R}$ bounded integrable such that $\|f(x, \theta) - f(x, \theta_0)\| \leq \|\theta - \theta_0\| g(x)$ for all $\theta \in N_0$, and
- (b) for some constants $c > 0$ and $k > 6/(p - 2)$ with $p > 4$, $\|f(x, \theta) - f(y, \theta)\| \leq c|x - y|^k$ for all $\theta \in \Theta$, on each piece S_i of their common support $S = \bigcup_{i=1}^m S_i \subset \mathbb{R}$, for some $m > 0$.

According to Park and Phillips (2001), all functions, which are bounded, piecewise smooth and integrable, are I – regular. k , which determines the smoothness of the function, depends on p that is determined by the existing moments of the underlying Data Generating Process (DGP). Hence, if $p < 8$, $k > 1$, which means the function should be constant.

⁴For example, $f\left(\frac{x_\epsilon}{\sqrt{T}}\right)$.

Definition 2.3. *Let*

$$f(\lambda x, \theta) = \nu(\lambda)h(x, \theta) + r(x, \lambda, \theta),$$

where ν is nonsingular. f is said to be H -regular if,

(a) h is regular on Θ , and

(b) $r(x, \lambda, \theta)$ is $o_p(\nu(\lambda))$, for all $\theta \in \Theta$.

ν and h are, respectively, called the **asymptotic order** and the **limit homogeneous function** of f .

Definition 2.3 requires conditions stronger than those in Park and Phillips (2001). Hence, the set of H -regular functions defined here is a subset of those defined by Park and Phillips (2001). In general, polynomial functions, logarithm functions and indicator functions all belong to H -regular class. The following tables summarize some popular I -regular or H -regular functions with respectively their integrals or limit homogeneous functions. It should be noticed that the definitions of I -regular or H -regular functions are based on certain general properties of the nonlinear function, which makes it almost impossible to be exhaustive about all existing examples.

Table 2.1: I -regular Functions

$f(x)$	$\int f(x)dx$
$\sin(x)$	$-\cos(x) + C$
$\cos(x)$	$\sin(x) + C$
$[\sin(x)]^2$	$\frac{x}{2} - \frac{1}{4}\sin(2x) + C$
$[\cos(x)]^2$	$\frac{x}{2} + \frac{1}{4}\sin(2x) + C$
$\frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$	$\frac{1}{2}[1 + \text{Erf}(\frac{(x-\mu)}{\sigma\sqrt{2}})] + C$

$C \in \mathbb{R}$ is a constant and Erf denotes the error function. Any probability density function which is bounded is I -regular. Moreover, if the random variable has at least the second moment being defined, the probability density function is square integrable.

Table 2.2: H – regular Functions

$f(\lambda x)$	$\nu(\lambda)$	$h(x)$	$r(x)$
$(\lambda x)^k, k > 0$	λ^k	x^k	0
$ \lambda x $	$ \lambda $	$ x $	0
$\text{Sign}(\lambda x), \lambda > 0$	1	$\text{Sign}(x)$	0
$\log \lambda x $	$\log \lambda $	1	$\log x $
$1\{\lambda x > a\}, \lambda > 0, a \in \mathbb{R}$	1	$1\{x > 0\}$	$1\{x > \frac{a}{\lambda}\} - 1\{x > 0\}$
$\Phi(\lambda x), \text{CDF}$	1	$1\{\lambda x > 0\}$	$\Phi(\lambda x)1\{\lambda x < 0\} + (\Phi(\lambda x) - 1)1\{\lambda x > 0\}$
$\max(\lambda x, a), \lambda > 0, a \in \mathbb{R}$	1	$\max(x, 0)$	$\max(x, \frac{a}{\lambda}) - \max(x, 0)$
$\min(\lambda x, a), \lambda > 0, a \in \mathbb{R}$	1	$\min(x, 0)$	$\min(x, \frac{a}{\lambda}) - \min(x, 0)$

The reciprocal function $\frac{1}{x}$ is not I – regular because it is not bounded and hence condition (a) of Definition 2.2 is not satisfied. $\frac{1}{x}$ is not H – regular either because it is not bounded at 0 which fails condition (b) of Definition 2.1 and condition (a) of Definition 2.3. Another concept on which the asymptotics of I – regular functions heavily rely is the Occupation Time Formula, which is extensively discussed by Revuz and Yor (1994) for continuous semimartingales. Chung and Williams (1990) shows that the formula also applies to the Brownian motion which is a special example of continuous semimartingales:

Lemma 2.1. (*Occupation Time Formula*) *Let f be locally integrable. B is a Brownian motion with the quadratic variation process $[B]$. Then,*

$$\int_0^t f(B(r))d[B]_r = \int_{-\infty}^{\infty} f(s)L_B(t, s)ds, \quad (2.1)$$

for all $t \in \mathbb{R}$.

For notational simplicity, the scaled local time L of B is defined by

$$L(t, s) = (1/\sigma_B^2)L_B(t, s). \quad (2.2)$$

Then, the following modification is made for Eq. (2.1),

$$\int_0^t f(B(r))dr = (1/\sigma_B^2) \int_{-\infty}^{\infty} f(s)L_B(t, s)ds = \int_{-\infty}^{\infty} f(s)L(t, s)ds, \quad (2.3)$$

because $d[B]_r = \sigma_B^2 dr$. The scaled local time $L(t, s)$ can therefore be regarded as the actual time spent by B up to time t in the neighborhood of s . Without loss of generality, let $\sigma_B^2 = 1$ and Lemma 2.1 can be represented as

$$\int_0^t f(W(r))dr = \int_{-\infty}^{\infty} f(s)L(t, s)ds, \quad (2.4)$$

where $W(r)$ is a standard Brownian motion. Applying $f(x) = 1\{|x - s| < \epsilon\}$ to Eq. (2.4) gives

$$\int_0^t 1\{|W(r) - s| < \epsilon\}dr = \int_{-\infty}^{\infty} 1\{|x - s| < \epsilon\}L(t, x)dx \approx \int_{-\epsilon}^{\epsilon} d(x - s)L(t, s), \quad (2.5)$$

for any s . Then,

$$L(t, s) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^t 1\{|W(r) - s| < \epsilon\}dr, \quad (2.6)$$

since $L(t, \cdot)$ is continuous on t . The above representation clearly explains why $L(t, s)$ is named local time of W at s . Later on, $L(t, 0)$, denoting the local time of a Brownian motion staying around point zero before time t , will be frequently used in this thesis.

2.2.2 Asymptotics for Sample Moments

In this section, the m_x -dimension nonstationary time series x_t considered is generated by

$$x_t = x_{t-1} + v_t, \quad (2.7)$$

where the innovation term v_t follows the linear process

$$v_t = \varphi(L)\varepsilon_t = \sum_{s=0}^{\infty} \varphi_s \varepsilon_{t-s}. \quad (2.8)$$

Furthermore, the 1-dimension stationary process w_t is defined as

$$w_t = \phi(L)\eta_t = \sum_{s=0}^{\infty} \phi_s \eta_{t-s}.$$

Let u_t be a stationary and ergodic martingale difference sequence. Then, following Park and Phillips (2001), we define $\xi_t = (u_t, \varepsilon'_{t+1}, \eta_{t+1})'$ and the filtration $\mathcal{F}_t = \sigma\{(\xi_s)_{-\infty}^{s=t}\}$, i.e., the σ -field generated by $(\xi_s)_{s \leq t}$. Then, the following assumptions are made for the innovation process ξ_t , and coefficients

φ_s and ϕ_s :

Assumption 2.1. *Assume*

- (a) (ξ_t, \mathcal{F}_t) is a stationary and ergodic martingale difference sequence,
- (b) $\mathbf{E}(\xi_t \xi_t' | \mathcal{F}_{t-1}) = \Sigma$,
- (c) $\sup_{t \geq 1} \mathbf{E}(\|\xi_t\|^q | \mathcal{F}_{t-1}) < \infty$ for some $q > 4$,
- (d) $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables,
- (e) $\mathbf{E}\|\varepsilon_t\|^p < \infty$ for some $p > 8$,
- (f) The distribution of ε_t is absolutely continuous with respect to the Lebesgue measure and has characteristic function $\Phi(\lambda)$ satisfying $\lim_{\|\lambda\| \rightarrow \infty} \Phi(\lambda) = o(\|\lambda\|^r)$ for some $r > 0$,
- (g) $\varphi(1)$ is nonsingular, $\sum_{s=0}^{\infty} s \|\varphi_s\| < \infty$,
- (h) $\sum_{s=0}^{\infty} s^{1/2} \|\phi_s\| < \infty$,

where Σ is partitioned as

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \Sigma_{u\varepsilon} & \sigma_{u\eta} \\ \Sigma_{\varepsilon u} & \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\eta} \\ \sigma_{\eta u} & \Sigma_{\eta\varepsilon} & \sigma_{\eta\eta} \end{bmatrix},$$

conformably with ξ_t .

The nonsingularity of $\varphi(1)$ implies that all integrated regressors are not linearly cointegrated. Moreover, let's assume d_t to be a m_d -dimension deterministic sequence which satisfies the following assumptions.

Assumption 2.2. *There exists a nonsingular sequence of normalizing matrices ν_{Td} such that if $d_T(r) = \nu_{Td}^{-1} d_{[Tr]}$ on the interval $[0, 1]$, then*

- (a) $\sup_{T \geq 1} \sup_{0 \leq r \leq 1} \|d_T(r)\| < \infty$,
- (b) $d_T \rightarrow_{L^2} d$ as $T \rightarrow \infty$ for some $d \in L^2[0, 1]$ such that $\int_0^1 d(r) d(r)' dr > 0$,

where $d_{[Tr]}$ is a continuous function of time.

Most conditions in Assumption 2.1 are commonly used in linear time series analysis and sufficient for the FCLT. The function $d_{[T^*]}$ in Assumption 2.2 is analog to the cadlag function defined for the partial sum of $I(1)$ processes. Assume that B_{x_i} , B_w and B_u are respectively the Brownian motions resulting from the weak convergence of the partial sum processes of x_{it} , w_t and u_t , for $i = 1, \dots, m_x$. Moreover, B_i , for $i = 1, \dots, m_x$, is another set of Brownian motions independent of B_{x_i} and B_u . The following lemma summarizes available asymptotic results for nonlinear transformations of $I(1)$ processes.

Lemma 2.2. *Suppose that x_t , u_t , w_t and d_t satisfy Assumptions 2.1 and 2.2, and a set of functions f_i , $g_i: \mathbb{R} \rightarrow \mathbb{R}^{k_i}$ for $i = 1, \dots, m_x$. Let f_i be I -regular and g_i be H -regular with the asymptotic order ν_i and the limit homogeneous function h_i . Assume that $\nu_{T_i} = \nu_i(\sqrt{T})$ is nonsingular, and h_i is piecewise differentiable with locally bounded derivative. Moreover, let \dot{h}_i the limit homogenous function of the first order derivative of g_i and continuous, and $\nu_{T_i} = T \dot{\nu}_{T_i}$ with $\dot{\nu}_{T_i}$ being the asymptotic order of the first order derivative. Define $g_{T_i} = \nu_{T_i}^{-1} g_i$ and $d_{T_i} = \nu_{T_i}^{-1} d_i$. Then, the following hold as $T \rightarrow \infty$:*

- (a) $\frac{1}{\sqrt{T}} \sum_{t=1}^T f_i(x_{it}) \xrightarrow{d} \int_{-\infty}^{\infty} f_i(s) ds L_i(1, 0).$
- (b) $\frac{1}{\sqrt{T}} \sum_{t=1}^T f_i(x_{it}) u_t \xrightarrow{d} \left(\int_{-\infty}^{\infty} f_i(s) f_i(s)' ds L_i(1, 0) \right)^{1/2} B_i(1).$
- (c) $\frac{1}{T^{3/4}} \sum_{t=1}^T f_i(x_{it}) w_t' \xrightarrow{p} 0.$
- (d) $\frac{1}{\sqrt{T}} \sum_{t=1}^T f_i(x_{it}) f_i(x_{it})' \xrightarrow{d} \int_{-\infty}^{\infty} f_i(s) f_i(s)' ds L_i(1, 0).$
- (e) $\frac{1}{\sqrt{T}} \sum_{t=1}^T f_i(x_{it}) f_j(x_{jt})' \xrightarrow{p} 0$ for $i \neq j.$
- (f) $\frac{1}{T} \sum_{t=1}^T g_{T_i}(x_{it}) \xrightarrow{d} \int_{-\infty}^{\infty} h_i(B_{x_i}(r)) dr.$
- (g) $\frac{1}{\sqrt{T}} \sum_{t=1}^T g_{T_i}(x_{it}) u_t \xrightarrow{d} \int_{-\infty}^{\infty} h_i(B_{x_i}(r)) dB_u(r).$
- (h) $\frac{1}{T} \sum_{t=1}^T g_{T_i}(x_{it}) w_t' \xrightarrow{p} 0.$
- (i) $\frac{1}{\sqrt{T}} \sum_{t=1}^T g_{T_i}(x_{it}) w_t' \xrightarrow{d} \int_{-\infty}^{\infty} h_i(B_{x_i}(r)) dB_w(r) + \int_{-\infty}^{\infty} \dot{h}_i(B_{x_i}(r)) dr \Lambda_{v_i w}$
with $\Lambda_{v_i w} = \sum_{s=0}^{\infty} E(v_{it} w_{t+s}).$
- (j) $\frac{1}{T} \sum_{t=1}^T g_{T_i}(x_{it}) g_{T_j}(x_{jt})' \xrightarrow{d} \int_0^1 h_i(B_{x_i}(r)) h_j(B_{x_j}(r))' dr.$
- (k) $\frac{1}{\sqrt{T}} \sum_{t=1}^T d_{T_i} f_i(x_{it}) = O_p(1).$
- (l) $\frac{1}{T} \sum_{t=1}^T d_{T_i} g_{T_i}(x_{it}) \xrightarrow{d} \int_{-\infty}^{\infty} d(r) h_i(B_{x_i}(r)) dr.$
- (m) $\frac{1}{\sqrt{T}} \sum_{t=1}^T f_i(x_{it}) g_{T_j}(x_{jt})' = O_p(1).$

The weak convergence in (a) (b) (d) (f) (g) (i) (j) and (l) holds jointly.

Remark 2.2.

- (1) $L_i(1, 0)$ denotes the local time of the Brownian motion B_{x_i} staying at point zero before time 1.
- (2) (i) follow from the results in Kasparis (2008). Moreover, the result could be obtained based on Saikkonen and Choi (2004) and De Jong (2005), which assume the DGP follows mixing or NED processes. The rest of Lemma 2.2 is based on Lemma 5 of Chang et al. (2001). The proofs can be found in Park and Phillips (1999), Park and Phillips (2001) and Chang et al. (2001). Ibragimov and Phillips (2008) synthesizes the results by using martingale convergence theorems.
- (3) As being pointed out by Chang et al. (2001), some interesting patterns are shown in the lemma. Parts (c) and (h) show that nonlinear functions of integrated processes are asymptotically orthogonal to stationary processes. Moreover, part (m) shows that the asymptotic orthogonality applies between I – *regular* functions of two integrated processes, regardless of how closely correlated these individual processes may be, if they are not linearly related.
- (4) All results in the lemma are based on functions defined on \mathbb{R} , namely $g_{Ti}(x_{it})$, but not $g_{Ti}(x_t)$ on \mathbb{R}^{m_x} . In fact, (f)-(j) and (l) still hold, if $g_{Ti}(x_{it})$ is replaced by $g_{Ti}(x_t)$, by extending the results in Christopeit (2009), Kasparis (2008) and Saikkonen and Choi (2004). However, for I – *regular* functions $f_i(x_{it})$, such an extension is not trivial given that the local time of a vector Brownian motion might degenerate to zeros.

2.2.3 Extensions

In this section, we categorize the work based on Lemma 2.2 into two branches: 1, extensions on the results in Lemma 2.2 to functions other than *regular*, I – *regular* or H – *regular* classes, i.e., *non – regular* functions of standardized random walks and exponential functions of random walks; 2, applications of the results in Lemma 2.2 to estimating nonlinear models and testing specifications.

Park and Phillips (1999) introduces an example of *non – regular* functions, $f(x) = \log|x|$, which is neither continuous at zero nor Riemann integrable, but locally integrable. Park and Phillips (1999) approximates the function with a sequence of regular functions indexed by the number of time series observations T :

$$f_T(x) = f(x)1\{|x - c| \geq c_T\} + f(c + c_T)1\{0 < x - c < c_T\} + f(c - c_T)1\{-c_T < x - c < 0\},$$

where $c = 0$ and c_T is a sequence of real numbers depending on T , for $f(x) = \log|x|$. Then, they obtain

$$\frac{1}{T} \sum_{t=1}^T f_T \left(\frac{x_{it}}{\sqrt{T}} \right) \xrightarrow{d} \int_0^1 f(B_{x_i}(r)) dr.$$

De Jong (2004) redefines the non-regular but locally integrable functions and establishes the above convergence result without using the approximating functions $f_T(x)$. But, De Jong (2004) requires the functions to have a finite number of discontinuity points and be monotone within the intervals where the function is continuous. A function that satisfies this monotone condition is defined as *monotone regular*. Moreover, Pötscher (2004) extended the results obtained by De Jong (2004) to a more general function class, locally integrable functions, by replacing the Riemann integrability with the Lebesgue integrability. In fact, *regular* functions, *non-regular* functions defined by Park and Phillips (1999) and *monotone regular* functions all belong to the class of locally integrable functions.

On the other hand, De Jong and Wang (2005) considers nonlinear functions of properly normalized $I(1)$ processes and nonintegrable function. *Proper normalization* means $f(\frac{x_{it}}{\sqrt{T^a}})$ for $0 \leq a < 1/2$ which encompasses the unnormalized case discussed by Park and Phillips (1999). De Jong and Wang (2005) requires f to be a continuous function, and bounded by an integrable and continuous function which is monotone on $(0, \infty)$ and $(\infty, 0)$. Being compared to the definitions used in Park and Phillips (1999), the Lipschitz condition is weakened to continuity and weaker restriction is required on the moments of the innovation terms. Nonintegrable functions such as $f(x) = |x|^{-m}$ for $m > 1$ are also considered by De Jong and Wang (2005). In addition, De Jong and Lee (2008) and De Jong (2009) obtain the asymptotic limiting results for exponential functions that are also introduced by Park and Phillips (1999). By taking a different perspective, Park (2003) concentrates on the asymptotic order of the approximation errors which is, as being mentioned in the paper, important for nonparametric and semiparametric methods, and bootstrap refinements.

In terms of model estimation, Park and Phillips (2001) and Chang et al. (2001) start with the following model

$$y_t = \pi' d_t + \alpha' w_t + \sum_{i \in \mathcal{I}} f_i(x_{it}; \beta_i) + \sum_{i \in \mathcal{H}} g_i(x_{it}; \beta_i) + u_t, \quad (2.9)$$

where \mathcal{I} denotes the set of *I-regular* functions and \mathcal{H} denotes the set of *H-regular* functions, and u_t satisfies Assumption 2.1. Moreover, d_t are the deterministic components, x_{it} are integrated regressors and w_t is a vector of linear processes. They obtained the asymptotic results for the Nonlinear Least Squares (NLS) estimator. The NLS estimates for the coefficients of the stationary

and deterministic regressors have the same convergence rates as those in standard linear regression models. Those of the integrated regressors are different from those in the linear regressions and depend on the type of functions. For $I - regular$ functions, the rate is $\sqrt[4]{T}$ and lower than that for the stationary regressors- \sqrt{T} . For $H - regular$ functions, the convergence rate depends on the asymptotic order $\sqrt{T}\dot{\nu}_T$. They could be faster or slower than \sqrt{T} depending on $\dot{\nu}_T$ being increasing or decreasing on T . Furthermore, they find that the asymptotic distributions for the estimated coefficients of the stationary components are independent from those of the rest of the model. The asymptotic distribution of the estimates of the $I - regular$ transformations are independent of those of the deterministic components and $H - regular$ transformations. The rate of consistency depends on both the stationarity property of the variables and the nonlinear function forms. Kasparis (2011) develops a residual based testing strategy to detect misspecified $H - regular$ functions. The Newey type and CUSUM type of tests are proved to be powerful against misspecified $H - regular$ functions. Kasparis (2008) discussed the consequence of misspecifying not only $H - regular$ functions but also $regular$ functions and $I - regular$ functions.

In contrast to the above mentioned literature that concentrates on general $I - regular$ or $H - regular$ functions, some other literature focuses on two popular models in empirical studies. Hong and Wagner (2008) and Hong and Wagner (2011) discuss estimation and inference issues on regressions with polynomial transformations of $I(1)$ processes and the application to the Environmental Kuznets Curve. On the other hand, Saikkonen and Choi (2004) studies the time series STR models with $I(1)$ regressors based on the extensions of the asymptotic results for $regular$ functions. Later, Choi and Saikkonen (2009) developed a subsample KPSS test to detect STR or polynomial regressions with $I(1)$ regressors. The test is shown to be applicable for both polynomial regressions and STR regressions.

2.3 Co-summability: Extension of Linear Cointegration

In the linear case, time series processes are conventionally classified as weakly or strictly stationary ($I(0)$), nonstationary ($I(1)$) and over-differenced ($I(-1)$). Furthermore, for $1/2 < d < 1$, the processes ($I(d)$) are named as long memory processes. The linear long run relationship among $I(1)$ (or $I(d)$) processes is defined as cointegration (fractional cointegration). In the econometric literature, there exists a variety of definitions of $I(0)$ processes. One of the most recent and complete definitions is probably introduced by Davidson (2009) which basically requires the standardized partial sum of the $I(0)$ process to satisfy a FCLT. Intuitively, in the nonlinear case, the term cointegration can

be generalized as nonlinear cointegration. However, the following example taken from Rico (2009) and Rico and Gonzalo (2011) shows that the definitions of $I(d)$ -ness and nonlinear cointegration are inadequacy when dealing with nonlinear transformations of $I(1)$ processes.

Example 2.1. *Square of a random walk: Assume that S_t is a random walk process, then*

$$S_t^2 = (\Delta S_t)^2 + 2S_{t-1}\Delta S_t + S_{t-1}^2.$$

If we consider the variance of the first difference of S_t^2 ,

$$Var[S_t^2 - S_{t-1}^2] = E[(\Delta S_t)^4] + (4t - 5)\sigma_{\Delta S}^4,$$

where $\sigma_{\Delta S}$ is the standard deviation of ΔS_t . As pointed out by Rico (2009) and Rico and Gonzalo (2011), ΔS_t^2 does not satisfy the definition of $I(0)$ by Davidson (2009) and any others in the literature. Hence, S_t^2 is not $I(1)$. Moreover, it is not even $I(d)$ for any $d > 0$ since the variance of $\Delta^d S_t^2$ always depends on t .

Apparently, we need a set of unified concepts to classify processes generated by different nonlinear transformations of $I(1)$ processes and interpret the relationship among such nonlinear transformations. To be general enough to encompass both linear and nonlinear transformations such as the one in Example 2.1, Rico (2009) and Rico and Gonzalo (2011) introduce the concepts of summability and co-summability which extend the definitions of integration and cointegration to the nonlinear situations. First, we introduce the definitions of *summability* and *co-summability*, and some examples introduced in Rico and Gonzalo (2011), which are used to interpret the nonlinear regression models with $I(1)$ regressors, such as polynomial regressions and STR regressions.

Definition 2.4. *A stochastic process S_t with positive variance is said to be summable of order d , represented as $S(d)$, if*

$$\bar{S}_T = \frac{1}{T^{1/2+d}} L(T) \sum_{t=1}^T (S_t - m_t) = O_p(1) \text{ as } T \rightarrow \infty$$

where d is the minimum real number that makes \bar{S}_T bounded in probability, m_t is a deterministic sequence, and $L(T)$ is a slowly-varying function.

The following example for Definition 2.4 is given in Rico and Gonzalo (2011).

Example 2.2. Assume that x_t is a random walk process as defined by Eq. 2.7, then

$$\bar{S}_T = \frac{1}{T^2 \sigma_v^2} \sum_{t=1}^T x_t^2 \Rightarrow \int_0^1 W^2(r) dr.$$

Hence, x_t^2 is $S(1.5)$ with, for instance, $L(T) = \sigma_v^2$, where σ_v^2 is the long run variance of v_t introduced in Eq. 2.7.

In the same paper, they show that summability can be used to check the balancedness of a nonlinear relationship and study nonlinear long run relationships.

Definition 2.5. A postulated relationship

$$y_t = f(x_t; \theta),$$

will be said to be balanced if y_t is $S(d_y)$, $z_t = f(x_t; \theta)$ is $S(d_z)$, and $d_y = d_z$.

Finally, after establishing the balancedness of a nonlinear model, Rico and Gonzalo (2011) use the concept of co-summability to analyze nonlinear long run relationships.

Definition 2.6. Two summable stochastic processes, y_t being $S(d_y)$ and x_t being $S(d_x)$, will be said to be co-summable if there exists $z_t = f(x_t; \theta)$ being $S(d_y)$ such that $u_t = y_t - f(x_t; \theta)$ is $S(d_u)$, with $d_u = d_y - \delta$ and $\delta > 0$. In short, $(y_t, z_t) \sim CS(d_y, \delta)$.

In this thesis, we modify Definition 2.4 to exclude $S(d)$ processes with $\bar{S}_T = o_p(1)$ and study a special co-summable relationship $CS(d_y, d_y)$ defined by Definition 2.6.

Definition 2.7. (Summability)⁵ A time series S_t is defined to be summable of degree d , symbolically represented as $S(d)$, if there exists a nonrandom sequence m_t such that

$$(a) \frac{1}{T^{1/2+d}} L(T) \sum_{t=1}^T (S_t - m_t) = O_p(1),$$

$$(b) \frac{1}{T^{1/2+d-\delta}} L(T) \sum_{t=1}^T (S_t - m_t) \text{ diverges to } \pm\infty \text{ for any } \delta > 0,$$

as $T \rightarrow \infty$, where $L(T)$ is a slowly-varying function at infinity.

⁵The sequence of m_t is not necessarily bounded or converging as Rico and Gonzalo (2011) pointed out, i.e., the time trend existing in a random walk with drifts.

Definition 2.8. (Strong Co-summability)⁶ For the time series y_t and x_t , the following regression model defines a strong co-summable relationship

$$y_t = g(x_t; \theta) + u_t,$$

if y_t and $g(x_t; \theta)$ are $S(d_y)$ with $d_y > 0$ and u_t is $S(0)$.

Remark 2.3.

- (1) First, we notice that Definition 2.7 excludes the following case: $\frac{1}{T^{1/2+d}}L(T)\sum_{t=1}^T(S_t - m_t) = o_p(1)$ and $\frac{T^\delta}{T^{1/2+d}}L(T)\sum_{t=1}^T(S_t - m_t)$ diverges to $\pm\infty$ for any $\delta > 0$. In Rico and Gonzalo (2011), this restriction is imposed by Assumption 2 in their paper.
- (2) As being pointed out by Rico and Gonzalo (2011), the definitions of $I(d)$ processes and cointegration in the linear case are encompassed by the above definitions respectively. Namely, $I(0) \subset S(0)$ and $I(d) \subset S(d)$ for any $d \geq 0$.
- (3) Definition 2.8 allows $f(x_t; \theta)$ to have additive components being $S(d_1)$ and $S(d_2)$ with $\min\{d_1, d_2\} \geq 0$ and $\max\{d_1, d_2\} = d_y$, for example, in the linear case, a model with both cointegrated $I(2)$ and $I(1)$ processes.
- (4) The requirement of u_t being $S(0)$ distinguishes the strong co-summability from weak co-summability defined in Rico and Gonzalo (2011). The weak co-summability only requires u_t to be $S(d_u)$ with $d_u > 0$ and $d_y > d_u$, where d_y is the degree of summability of y_t . The linear cointegration case is an example of the strong co-summability relationship with u_t being $I(0)$. Moreover, two cointegrated $I(2)$ processes with $I(1)$ residuals define a weak co-summability relationship of which the detailed discussion can be found in Haldrup (1998).
- (5) It is worth noticing that the models considered in this thesis actually define a special case of strong co-summable relationship where the regressors are generated by nonlinear transformations of $I(1) \subset S(1)$ processes and the regression error term is $I(0) \subset S(0)$.⁷ Moreover, we further distinguish between simple strong co-summability and multiple strong co-summability in the later chapters.

⁶It is worth of noticing that the u_t defined in Definition 2.8 is more general than the u_t defined in Assumption 2.1. Moreover, u_t will be repeatably used in this thesis to denote the regression error terms. Hence, it will be defined by the different assumptions in different chapters but its properties should be clear under the relevant context.

⁷As being pointed out by Rico (2009) and Rico and Gonzalo (2011), an over-differenced process is $S(0)$ but not $I(0)$.

2.4 Nonlinear Asymptotic Theory for Stationary Processes

Given our interest in co-summable relationship defined by models with both nonlinearly into variables and parameters, the asymptotic results of the NLS estimator require the analysis to be different from that of the Ordinary Least Squares (OLS) estimators which are defined as the close-form solution of the objective function in the linear case. Conventionally, the consistency and asymptotic normality of the NLS estimator are treated separately in the econometric literature. The consistency results of NLS estimators are based on the asymptotic properties of the objective function, which is different from that of OLS estimators. Moreover, the positive conclusion of consistency is prerequisite to the asymptotic distribution of NLS estimators. Provided that the NLS estimator is a special case of \mathbf{M} -estimators, this section introduces the methods designed, in the nonlinear econometric literature, for the more general \mathbf{M} -estimators when the regressors are $I(0)$.

2.4.1 Consistency

The classical asymptotic theory for nonlinear regression models is obtained for stationary and ergodic processes. Hence, in this section, we assume the m_x -dimension vector of the time series x_t to be $I(0)$, i.e., martingale difference processes, mixing processes or NED process. The existence of \mathbf{M} -estimators requires that the objective function must be continuous and measurable on a compact parameter space. The consistency proof, introduced by Jennrich (1969), consists of two main ingredients: uniform convergence over the parameter space and the identifiable uniqueness of the true parameter values. This uniform convergence boils down to proving uniform LLNs.

For any observed sample with size T , $Q_T(x_1, \dots, x_T; \gamma, \theta)$ is a real valued function defined on $(\mathbb{R}^{m_x})^T \times \Gamma \times \Theta$. θ is the vector of parameters of interest and γ is the vector of possible nuisance parameters. We assume further that $Q_T(x_1, \dots, x_T; \gamma, \theta)$ is measurable on the sigma field generated by the probability space Ω of $\{x_t\}_{t=-\infty}^{\infty}$, for all $(\gamma, \theta) \in \Gamma \times \Theta$. Then, the \mathbf{M} -estimators $\hat{\theta}_T$ for the parameters of interest θ , corresponding to the objective function Q_T , are defined as

$$\hat{\theta}_T = \arg \min_{\theta \in \Theta} Q_T(x_1, \dots, x_T; \hat{\gamma}_T, \theta), \quad (2.10)$$

where Q_T are generally some “distance” functions and $\hat{\gamma}_T$ are some consistent estimators of γ . Furthermore, (Γ, ρ_Γ) and (Θ, ρ_Θ) are assumed to be compact non-empty metric spaces with respectively the metric ρ_Γ and ρ_Θ . γ can be viewed as a set of nuisance parameters which increase the flexibility of the model specification. The true model is indexed by γ_{T0} and θ_0 which are unknown. The esti-

mation problem is then to infer the true values of the parameters. The assumption of γ_{T0} depending on T encompasses the cases where deterministic time trends present in the model. The following lemma based on Pötscher and Prucha (1991a) shows the convergence behavior of \mathbf{M} -estimators for very general nonlinear models.

Lemma 2.3. *Given sequences of functions $Q_T : \Omega \times \Gamma \times \Theta \rightarrow \mathbb{R}^l$, $\bar{Q}_T : \Gamma \times \Theta \rightarrow \mathbb{R}^l$, for $l \in \mathbb{N}$, $\hat{\gamma}_T : \Omega \rightarrow \Gamma$ and a nonstochastic sequence $\bar{\gamma}_T \in \Gamma$, let $R_T(\omega, \theta) = Q_T(x_1, \dots, x_T; \hat{\gamma}_T, \theta)$ and $\bar{R}_T(\theta) = \bar{Q}_T(\bar{\gamma}_T, \theta)$. Let $\{\bar{Q}_T : T \in \mathbb{N}\}$ be uniformly equicontinuous on $\Omega \times \Theta$.*

(a) *If $\rho_\Gamma(\hat{\gamma}_T, \bar{\gamma}_T) \rightarrow 0$ a.s. [i.p.], as $T \rightarrow \infty$, and if a.s. [i.p.]*

$$\sup_{\Theta \times \Gamma} |Q_T(x_1, \dots, x_T; \gamma, \theta) - \bar{Q}_T(\gamma, \theta)| \rightarrow 0, \text{ as } T \rightarrow \infty, \quad (2.11)$$

then a.s. [i.p.]

$$\sup_{\Theta} |R_T(\omega, \theta) - \bar{R}_T(\theta)| \rightarrow 0, \text{ as } T \rightarrow \infty. \quad (2.12)$$

(b) *The family $\{\bar{R}_T : T \in \mathbb{N}\}$ is uniformly equicontinuous on Θ , with \mathbb{N} denoting the set of positive integers.*

Let θ_{T0} be an identifiably unique sequence of minimizers of $\bar{Q}_T(\theta)$, then for any sequence $\hat{\theta}_T$ such that eventually

$$Q_T(\omega, \hat{\gamma}_T, \hat{\theta}_T) = \inf_{\Theta} Q_T(\omega, \hat{\gamma}_T, \theta) \quad (2.13)$$

holds, we have

$$\rho_\Theta(\hat{\theta}_T, \theta_{T0}) \rightarrow 0$$

a.s. [i.p.], as $T \rightarrow \infty$, where a.s. and i.p. stand for, respectively, “almost surely” and “in probability” with respect to the probability measure defined on Ω .

Remark 2.4.

(1) The above lemma only summarizes conditions which are sufficient for the consistency of M -estimators.

For more detailed treatment of issues such as the existence of the estimator, uniform LLNs and identifiable uniqueness, the reader is referred to White (1980), Domowitz and White (1982), Pötscher and Prucha (1997), Pötscher and Prucha (1991a) and Pötscher and Prucha (1991b).

- (2) $\theta_{T0} = \theta_0$, which is commonly assumed in the econometric literature, a special case resulting from Lemma 2.3.
- (3) Generally speaking, Lemma 2.3 shows that three conditions suffice the consistency of the estimator: compactness of the parameter space, identifiable uniqueness of the true values of the parameters, and equicontinuity and uniform convergence of the objective function over the parameter spaces. The second condition is usually assumed in the nonlinear econometric literature and it can be verified for some models, i.e., models linear in the parameters. The first condition can also be verified for some cases to which the models introduced in the following chapters belong. The third condition follows from uniform LLNs in the stationary cases. For the nonstationary case, it can be verified by assuming or proving that the objective function is equicontinuous on the parameter spaces, which is indicated by Definitions 2.1, 2.2 and 2.3. Besides, the convergence in probability or almost sure only holds for the nonstationary nonlinear panel models. For nonstationary nonlinear time series regressions, only convergence in distribution can be obtained.

2.4.2 Asymptotic Distribution

The standard approach for deriving the asymptotic distribution is based on the score vectors and the Hessian matrix of the objective function. Hence, the objective function generally needs to be twice continuously differentiable with respect to the parameters and the true values of the parameters must be in the interior of the parameter space. The convergence of the score vectors and the Hessian matrix follows from the consistency results obtained from the previous section, uniform LLNs and CLT in the stationary cases.

In order to obtain the score vectors and Hessian matrix, the first order derivatives of the objective function $Q_T(x_1, \dots, x_T; \hat{\gamma}_T, \theta)$ with respect to θ are linearized based on the Taylor expansion. Let $\dot{Q}_{T\theta}$, $\ddot{Q}_{T\theta\theta}$ and $\ddot{Q}_{T\theta\gamma}$ be the first and second derivatives of Q_T with respect to θ and γ . The asymptotic distribution of $\hat{\theta}_T$ can be obtained by the first order Taylor expansion of $\dot{Q}_{T\theta}$, if we assume Γ to be an empty space,

$$\dot{Q}_{T\theta}(x_1, \dots, x_T; \hat{\theta}_T) = \dot{Q}_{T\theta}(x_1, \dots, x_T; \theta_0) + \ddot{Q}_{T\theta\theta}(x_1, \dots, x_T; \tilde{\theta}_T)(\hat{\theta}_T - \theta_0),$$

where $\tilde{\theta}_T$ lies in the line segment connecting $\hat{\theta}_T$ and θ_0 in case θ is a scalar. The subsequent assumptions provide the set of sufficient conditions of deriving the asymptotic distribution of $\hat{\theta}_T$. M_T and

N_T are two normalizing sequences. In the stationary case, Pötscher and Prucha (1991b) formulated the conditions for asymptotic normality in a rather general fashion as follows:

Assumption 2.3. *The follows are assumed:*

- (a) *The parameter spaces Γ and Θ are measurable subsets of Euclidean space \mathbb{R}^{p_γ} and \mathbb{R}^{p_θ} , respectively.*
- (b) *$Q_T(x_1, \dots, x_T; \gamma, \theta)$ is measurable with respect to the sigma-field generated by Ω for all $(\gamma, \theta) \in \Gamma \times \Theta$ and $Q_T(x_1, \dots, x_T; \cdot)$ is a.s. twice continuously and partially differentiable at every point (γ, θ) in the interior of $\Gamma \times \Theta$.*
- (c) *The estimators $(\hat{\gamma}_T, \hat{\theta}_T)$ take their values in $\Gamma \times \Theta$. There exists a non-random sequence $(\gamma_{T0}, \theta_{T0}) \in \Gamma \times \Theta$, which is eventually uniformly in the interior of $\Gamma \times \Theta$, such that $\hat{\theta}_T - \theta_{T0} = o_p(1)$, $\hat{\gamma}_T - \gamma_{T0} = o_p(1)$ and $M_T(\hat{\gamma}_T - \gamma_{T0}) = O_p(1)$ for a sequence of square matrices (possibly random) M_T , which are non-singular with probability tending to one.*
- (d) *The sequence $\hat{\theta}_T$ satisfies,*

$$TN_T^{+'} \dot{Q}_{T_{\theta'}}(x_1, \dots, x_T; \hat{\gamma}_T, \hat{\theta}_T) = o_p(1)$$

for a sequence of square matrices (possibly random) N_T , which are non-singular with probability tending to one.

- (e) *For all sequences of random vectors $(\tilde{\gamma}_T, \tilde{\theta}_T)$ with $\tilde{\gamma}_T - \gamma_{T0} = o_p(1)$ and $\tilde{\theta}_T - \theta_{T0} = o_p(1)$ we have*

$$TN_T^{+'} \ddot{Q}_{T_{\theta\theta}}(x_1, \dots, x_T; \tilde{\gamma}_T, \tilde{\theta}_T) N_T^+ - C_T = o_p(1),$$

for a sequence of (possibly random) matrices C_T , which are non-singular with probability tending to one and satisfy $\|C_T\| = O_p(1)$ and $\|C_T^+\| = O_p(1)$.

- (f) *For all sequences $(\tilde{\gamma}_T, \tilde{\theta}_T)$ as in (e) we have*

$$TN_T^{+'} \ddot{Q}_{T_{\theta\gamma}}(x_1, \dots, x_T; \tilde{\gamma}_T, \tilde{\theta}_T) M_T^+ = o_p(1).$$

- (g) *There exists a sequence of (possibly random) matrices D_T with $\|D_T\| = O_p(1)$, such that,*

$$-TN_T^{+'} \dot{Q}_{T_\theta}(x_1, \dots, x_T; \tilde{\gamma}_T, \tilde{\theta}_T) = D_T \xi_T + o_p(1),$$

where ξ_T and ξ are vectors satisfying $\xi_T \xrightarrow{d} \xi$.

The assumption of $\hat{\theta}_T - \theta_{T0} = o_p(1)$ and $\hat{\gamma}_T - \gamma_{T0} = o_p(1)$ in Condition (c) follows from consistency results in Lemma 2.3. In the special case where there is no nuisance parameter γ and $\bar{\theta}_T \equiv \theta_0$ independent on T , the condition that $(\bar{\gamma}_T, \bar{\theta}_T)$ lies uniformly in the interior of $\Gamma \times \Theta$ reduces to the condition that (γ_0, θ_0) lies in the interior of $\Gamma \times \Theta$. Condition (d) is actually satisfied by the definition of the \mathbf{M} -estimators. Namely, if $\hat{\theta}_T$ is an interior minimizer of $Q_T(x_1, \dots, x_T; \hat{\gamma}_T, \theta)$ or if $\hat{\theta}_T$ is a solution of the first order condition, then Assumption 2.3 (d) is trivially satisfied for any choice of N_T . Condition (e) assumes the Hessian matrices converge in probability to constant matrices or random matrices. $\tilde{\gamma}_T$ and $\tilde{\theta}_T$ are obtained from the Taylor series expansion and lie in the parameter space where the asymptotic normality is analyzed. As $N_T(\hat{\theta}_T - \theta_{T0})$ is asymptotically equivalent to a linear function of the score $\dot{Q}_{T\theta'}(x_1, \dots, x_T; \tilde{\gamma}_T, \theta_{T0})$ and of $\hat{\gamma}_T - \gamma_{T0}$, the asymptotic distribution of $\hat{\theta}_T$ depends on the asymptotic behavior of the score vector as well as that of $\hat{\gamma}_T$. Hence, Condition (f), in conjunction with $M_T(\hat{\gamma}_T - \gamma_{T0}) = O_p(1)$, is necessary to eliminate this dependence on the asymptotic distribution of $\hat{\gamma}_T$, which ensures that γ is a nuisance parameter. In practice, both Condition (e) and (f) boil down to the LLNs. Condition (g), in the stationary case, is obtained from CLT and ξ is normally distributed. In addition, the normalizing sequences will be of the form $M_T = T^{1/2}\mathbf{I}$ and $N_T = T^{1/2}\mathbf{I}$, C_T and D_T will be non-random. Asymptotically, non-randomness of C_T , N_T and M_T implies that $C_T^+ = C_T^{-1}$, $N_T^+ = N_T^{-1}$ and $M_T^+ = M_T^{-1}$ for large T .

Based on Assumption 2.3, the following lemma given in Pötscher and Prucha (1991b) provides the asymptotic distribution of $\hat{\theta}_T - \theta_{T0}$.

Lemma 2.4. *Given Assumption 2.3 holds, then,*

$$N_T(\hat{\theta}_T - \theta_{T0}) = C_T^+ D_T \xi_T + o_p(1),$$

where $\xi_T \xrightarrow{d} \xi$. Furthermore, if $C_T^+ D_T \rightarrow A$ i.p. and A is non-random, then,

$$N_T(\hat{\theta}_T - \bar{\theta}_T) \xrightarrow{d} A\xi.$$

More generally, for any sequence of (possibly random) matrices G_T with $\|G_T\| = O_p(1)$, we have

$$G_T N_T(\hat{\theta}_T - \theta_{T0}) = G_T C_T^+ D_T \xi_T + o_p(1),$$

where $\xi_T \xrightarrow{d} \xi$. If furthermore $G_T C_T^+ D_T \rightarrow A^*$ i.p. and A^* is non-random, then,

$$G_T N_T(\hat{\theta}_T - \theta_{T0}) \xrightarrow{d} A^* \xi.$$

This lemma uses the distribution function of $A\xi$ as an approximation to that of $N_T(\hat{\theta}_T - \bar{\theta}_T)$. The key point to justify the approximation is establishing the converging limit A of $C_T^+ D_T$. Then, in such a case, the lemma implies that the cumulative distribution function of $N_T(\hat{\theta}_T - \bar{\theta}_T)$ converges to the cumulative distribution function of $A\xi$. If $C_T^+ B_T$ does not converge, it is less obvious in which sense the distribution of $N_T(\hat{\theta}_T - \bar{\theta}_T)$ and of $C_T^+ B_T$ are “close” to each other. In practice, if C_T and D_T can be chosen to be non-random, the distributional convergence of $C_T^{-1} D_T \xi_T$ to $C_T^{-1} D_T \xi$ is a consequence of Skorohod’s Representation Theorem. We need to find random variables $\tilde{\xi}_T$ and $\tilde{\xi}$ defined on some probability space such that $\xi_T =_d \tilde{\xi}_T$ and $\tilde{\xi} =_d \xi$ and $\tilde{\xi}_T \rightarrow \tilde{\xi}$ a.s..

2.5 Conclusion

In this chapter, we introduce the main tools which are needed for analyzing the asymptotic properties of NLS estimators for a panel strong co-summable regression model with large n and large T . The tools include existing asymptotic results for nonlinear transformations of I(1) processes, summability and co-summability, and nonlinear asymptotic theory for stationary time series regression models. It seems that a compact parameter space and twice differentiability of the objective function with respect to the parameters are indispensable. Besides, assumptions on the identifiability of the true values of the parameters are sometimes also necessary. The task requiring further efforts is deriving uniform convergence. On the other hand, some interesting issues and questions are raised, which leads to the following conclusive remarks.

- In the co-summable regressions defined by linear cointegration with short run dynamics, the 2-stage Least Squares estimator has been shown to have the same rate of convergence in probability as the normal 1-step Least Squares estimator. However, this is not the case where the co-summability is defined by different nonlinear transformations of I(1) regressors. For example, we can show both theoretically and by simulation that for the STR co-summability with I(1) regressors, the 2-stage NLS estimator for the parameters of the linear components has the rate of consistency being \sqrt{T} . But, the rate is T for the 1-step NLS estimator.

- For nonlinear stationary cases, the identification of the true values of the parameters generally depends on the function form because existing uniform LLNs leads to deterministic limit of the objective function. While for the nonlinear nonstationary cases, only results similar to continuous mapping theory are available and the asymptotic limit of the objective function is generally a random variable. Hence, the identification condition is assumed sometimes but not all can be verifiable in the time series co-summable regression model. This condition can be generally verified if we have cross-section independence in the panel co-summable regressions.
- The asymptotic results for $H - regular$ functions with endogenous regressors have been obtained, but the issue is still open and the solution is not clear for $I - regular$ functions.
- Spurious regression: Given that the concepts of integration and cointegration are extended to summability and co-summability for possible nonlinearity, the issue of spurious regressions is worth of a reconsideration. For example, if z_t and x_t are $I(1)$ scalar processes. Regressing z_t upon $f(x_t)$, we have the following expression,

$$\frac{\sum_t z_t f(x_t)}{\sum_t f^2(x_t)}.$$

If f is assumed to be $I - regular$, the above expression diverges to infinity as $T \rightarrow \infty$. Consequently, the spurious relation between z_t and $f(x_t)$ may always be statistically significant. On the other hand,

$$\frac{\sum_t 1\{z_t > 0\}1\{x_t > 0\}}{\sum_t 1\{x_t > 0\}^2}$$

converges to

$$\frac{\int_0^1 1\{B_z(r) > 0\}1\{B_x(r) > 0\}dr}{\int_0^1 1\{B_x(r)\}dr},$$

which is similar to the spurious regressions in the linear case.

- On the issues of estimation and inference, Park and Phillips (2001) considers the NLS estimator for the following strong co-summable model,

$$y_t = f(x_t, \theta) + u_t,$$

where x_t is a scalar $I(1)$ process and u_t is a martingale difference process. f is assumed to be either $I - regular$ or $H - regular$. They base the analysis and derivation on Lemmas 2.3 and 2.4. Moreover, they assume that all derivatives and f belong to the same class of

functions. On the other hand, Chang et al. (2001) considers a more general model specification (2.9). However, they base their analysis on the Wooldridge (1994) that does not consider the consistency issue with the objective function. As De Jong and Hu (2010) pointed out that the approach of Wooldridge (1994) may lead to non-identification of the true values of the parameters in the nonlinear nonstationary case. Considering the STR co-summable models, the time series results of Park and Phillips (2001) cannot be applied directly because some of the first order derivatives of the logistic smooth transition function are I -regular while the logistic smooth transition function is H -regular. On the other hand, we are cautious about applying the results obtained in Chang et al. (2001) due to the discussion of De Jong and Hu (2010). Hence, we need to extend the method used by Park and Phillips (2001) to the models considered by Chang et al. (2001), in order to derive sequential asymptotics for STR co-summable models. Chang and Park (2003) provides asymptotic results for co-summability defined by general index models which encompass STR co-summable models as a special case. However, they base their derivation on Wooldridge (1994). Moreover, Saikkonen and Choi (2004) also study the STR co-summability. They propose to approximate the STR co-summable model with a sequence of STR co-summable models where the $I(1)$ regressors are scaled by \sqrt{T} . Namely, they consider

$$f\left(\frac{\sqrt{T_0}}{\sqrt{T}}x_t\right),$$

with T_0 be the size of the current sample. The asymptotic limit depends on $T \rightarrow \infty$. Then, the original model is obtained when $T = T_0$. According to Definitions 2.1 and 2.2, $f(\frac{\sqrt{T_0}}{\sqrt{T}}x_t)$ is *regular* and $f(x_t)$ is I -regular. In fact, they imply different orders of summability with respect to the dependent variables.

- In the following chapters of this thesis, the attention is paid to the panel data sets with both large n cross section dimension and large T time series dimension. However, the possibility of having either small n or small T panels should not be ruled out in practice. For the case where n becomes large and T is relatively small, the cross sectional asymptotic results, where $n \rightarrow \infty$ while T is finite, are expected to be the better approximation for finite sample inference than the large n and large T asymptotic results. Accordingly, due to the finiteness of T , the time series features of the data are very hard to be consistently estimated. Hence, it might not be important to distinguish non-stationarity with stationarity in this case. On the other hand, if $T \rightarrow \infty$ while n is finite, the panel can be treated as multiple time series. In this case, the data

contains sufficient information on their time series features such as non-stationarity. Therefore, large T asymptotic results are expected to be the better approximation to the finite sample inference under this situation. Nevertheless, neither case is the focus of this thesis and they demand further studies and research.

Pooled Nonlinear Least Squares Estimation for Independent Co-summable Panels

3.1 Introduction

In econometric applications, nonlinear parametric models have been considered as they offer more flexibility than linear models in describing the data, for example, in modeling multiple equilibria and business cycle behaviors. Popular nonlinear models in the literature include Threshold Regression models, Smooth Transition Regression models and Markov-Switching models. Moreover, the wide availability of well constructed and maintained panel data sets with a large number of time series observations (T) and a large number of cross-section units (n) inspire further applications of nonlinear parametric models on panel data analysis. Recent developments include inter alia Baum et al. (2001), Lo and Zivot (2001), Taylor (2001) and González et al. (2005). Most of the applications found in the literature are about macroeconomic analysis where the time series components are generally considered to be nonstationary.

In this chapter, we apply the asymptotic results obtained by Park and Phillips (1999), Park and Phillips (2001) and Chang et al. (2001) on nonlinear asymptotics for $I(1)$ regressors to extend the analysis in Phillips and Moon (1999b) on linear nonstationary panel data models to nonlinear nonstationary cross-sectionally independent panel data models. In addition, the concepts of summability and co-summability proposed by Rico (2009) and Rico and Gonzalo (2011) are applied and modified to interpret the nonlinear long run relationship defined by nonlinear regression models with $I(1)$ regressors, which are viewed as the extension of integration and cointegration. The focus of this chapter is to develop the asymptotic results for the Pooled Nonlinear Least Squares (NLS) estimator because pooling is one of the major advantages considered by the literature for panel data models, which could provide extra information other than single time series on the commonality in nonlinear and trending behaviors. In order to fulfill the task, we start with extending the time series asymptotics in Park and Phillips (2001) to the product of I -regular functions and H -regular functions.

Then, we provide a set of sufficient conditions for the existence of the Pooled NLS estimator. Moreover, the consistency property and asymptotic distribution are derived for $T \rightarrow \infty$ being followed by $n \rightarrow \infty$. Finally, the results for the panel Logistic Smooth Transition Co-summable Regression (LSTCR) models are illustrated by the simulation study with a robust check on different estimators for the covariance matrix.

The rest of the chapter is organized as follows. Section 3.2 introduces the basic model and lays out assumptions. In Section 3.3, the preliminary results are developed with the discussion of the existence of the Pooled NLS estimator. Furthermore, the asymptotic results are derived for simple strong co-summable regression models including *I-regular* and *H-regular* functions. Section 3.4 contains the consistency and asymptotic distribution results for multiple strong co-summable regression models. Section 3.5 takes the LSTCR model as an example, and proposes several estimators for the covariance matrix. In Section 3.6, the simulation study is conducted based on the LSTCR model. Moreover, the performance of different estimators for the covariance matrix is compared for inference purposes. Some concluding remarks are made in Section 3.7.

A word about the notations. For a vector $x = (x_i)$ or a matrix $A = (a_{ij})$, the modulus $|\cdot|$ is taken element by element. Therefore, $|x| = (|x_i|)$ and $|A| = (|a_{ij}|)$. The standard Euclidean norm of a vector is denoted by $\|\cdot\|$, i.e. $\|x\|^2 = \sum_i (x_i)^2$. For a matrix, $\|A\|$ signifies the operator norm defined by $\|A\| = \sup_x \|Ax\|/\|x\|$. The maximum of the moduli is denoted by $\overline{\|\cdot\|}$, i.e., $\overline{\|x\|} = \max_i |x_i|$ and $\overline{\|A\|} = \max_{i,j} |a_{ij}|$. Moreover, the k by k dimension identity matrix is denoted by \mathbf{I}_k . For a function, which can be vector- or matrix-valued, $\overline{\|\cdot\|}_K$ signifies the supremum over a subset K of its domain, so that $\overline{\|f\|}_K = \sup_{x \in K} \|f(x)\|$. The subset K , over which the supremum is taken, will not be specified if it is clear from the context. Furthermore, the indicator function is written as $1\{\cdot\}$. Letters such as p, q, m, l and δ generally denote positive integers and the definitions should be clear under the context. $=_d$ denotes equality in distribution. In addition, standard terminologies and notations in probability and measure theory are used throughout the rest of the chapter such as almost sure convergence ($\xrightarrow{a.s.}$), convergence in probability (\xrightarrow{p}), convergence in distribution (\xrightarrow{d}) and weak convergence (\Rightarrow). $W(r)$ denotes a vector of standard Brownian motions and $B(r)$ denotes a vector of general Brownian motions. Their dimensions are clear from the context. For a function $f(x, \theta)$, \dot{f} , \ddot{f} and \dddot{f} denote, respectively, the first, second and third derivatives of f with respect to θ . They are possibly vectors and arranged by the lexicographic ordering of their indices based on the vector θ .

3.2 The Model and Assumptions

In contrast to a large part of the current literature on large (T) and large (n) panels that almost exclusively deals with linear models, we impose homogeneity of the parameter θ and assume cross-sectional independence in the models considered in this chapter. Therefore, we consider the following panel nonlinear regression model for a scalar endogenous variable $y_{i,t}$:

$$y_{i,t} = f(z_{i,t}, \theta)x_{i,t} + u_{i,t}, \quad (3.1)$$

where θ is a finite dimension vector of parameters, $z_{i,t}$ and $x_{i,t}$ are scalar regressors assumed to be generated by

$$\begin{bmatrix} x_{i,t} \\ z_{i,t} \end{bmatrix} = \begin{bmatrix} x_{i,t-1} \\ z_{i,t-1} \end{bmatrix} + \begin{bmatrix} v_{x,i,t} \\ v_{z,i,t} \end{bmatrix}, \quad (3.2)$$

with common initialization at $t = 0$ satisfying

$$(z_{i,0}, x_{i,0})' \text{ being IID across } i \text{ with } \mathbf{E}\|(z_{i,0}, x_{i,0})'\|^4 < \infty.$$

The innovation term $V_{i,t} = (v_{x,i,t}, v_{z,i,t})'$ is assumed to be generated by the linear process with random coefficients,

$$V_{i,t} = \varphi_i(L)\eta_{i,t} = \sum_{s=0}^{\infty} \varphi_{i,s} \begin{bmatrix} \varepsilon_{x,i,t-s} \\ \varepsilon_{z,i,t-s} \end{bmatrix}, \quad (3.3)$$

where $\varphi_{i,s}$ are 2 by 2 matrices. The functions $f(\cdot, \theta)$ is assumed to be all known but not θ and identical across units. Besides, the parameters to be estimated, θ , are also identical across units. Model (3.1) can be viewed as the extension of the homogeneous panel linear cointegration defined in Phillips and Moon (1999b). There are two main reasons for considering the specification in (3.1). First, most of the popular nonlinear models such as the Threshold models and the Smooth Transition models have their simple form been encompassed by (3.1). Second, according to the author's knowledge, there is no technical tools in the existing literature to facilitate the derivation of $\lim_{T \rightarrow \infty} \sum_{t=1}^T f(x_{i,t}, z_{i,t}, \theta)$ with f being $I - regular$ and the variables being $I(1)$. The results of the previous mentioned limit depends on the occupation time at zero of a vector Brownian motion $(x_{i,t}, z_{i,t})'$. This occupation time is zero if the vector Brownian motion is transient $\varepsilon_{i,t}$.

Following Park and Phillips (2001), we define $\xi_{i,t} = (u_{i,t}, \eta'_{i,t+1})'$ and the filtration $\mathcal{F}_{i,t} = \sigma(\{\xi_{i,s}\}_{s=-\infty}^{s=t})$, i.e., the σ -field generated by $\{\xi_{i,s}\}_{s \leq t}$. The following assumption is made for the innovation term $\xi_{i,t}$.

Assumption 3.1. (DGP)

- (a) $(\xi_{i,t}, \mathcal{F}_{i,t})$ is a stationary and ergodic martingale difference sequence, which is independent across i ,
- (b) the vectors $\eta_{i,t}$ are IID for all i and t with $\mathbf{E}(\eta_{i,t}) = 0$ and $\mathbf{E}(\eta_{i,t}\eta'_{i,t}) = \mathbf{I}_2$,
- (c) $\mathbf{E}(\xi_{i,t}\xi'_{i,t} | \mathcal{F}_{i,t-1}) = \Sigma_i$, where Σ_i is positive definite and nonsingular, which is partitioned as

$$\Sigma_i = \begin{bmatrix} \sigma_{uu,i}^2 & \sigma_{u\varepsilon_x,i} & \sigma_{u\varepsilon_z,i} \\ \sigma_{\varepsilon_x u,i} & 1 & 0 \\ \sigma_{\varepsilon_z u,i} & 0 & 1 \end{bmatrix}, \quad (3.4)$$

- (d) $\mathbf{E}\|\eta_{i,t}\|^p < \infty$ for some $p > 8$,
- (e) $\sup_{t \geq 1} \mathbf{E}(\|\xi_{i,t}\|^q | \mathcal{F}_{i,t-1}) < \infty$ for some $q > 4$,
- (f) The distribution of $\eta_{i,t}$ is absolutely continuous with respect to the Lebesgue measure and has characteristic function $\Phi(\lambda)$ satisfying

$$\lim_{\|\lambda\| \rightarrow \infty} \Phi(\lambda) = o(\|\lambda\|^\delta),$$

for some $\delta > 0$.

Conditions (a) and (c)-(f) are required by exploring the time dimension properties as $T \rightarrow \infty$, which is based on the nonlinear time series regression asymptotic results developed in Chang et al. (2001). The error terms $u_{i,t}$ are assumed to be martingale difference processes because the derivation of the limit of $\sum_{t=1}^T f(z_{i,t}, \theta)x_{i,t}u_{i,t}$ is based on the convergence theorems of continuous martingale difference processes. Moreover, condition (f) is also due to the existing method of deriving the limit of the local time of a random walk process. It is not clear how these assumptions could be relaxed if no better way of derivation can be discovered. Condition (b) and the restriction on Σ_i is conventional as we assume the data generating processes of $x_{i,t}$ and $z_{i,t}$ follow a linear process, e.g., Phillips and Moon (1999b).

Assumption 3.1'. *Items (a) to (f) from Assumption 3.1 hold with*

$$\Sigma_i = \begin{bmatrix} \sigma_{uu,i}^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.5)$$

This assumption will be needed for the case where $f(z_{i,t}, \theta)$ is asymptotically homogeneous functions of $z_{i,t}$. Moreover, a set of assumptions regarding the random coefficients $\varphi_{i,s}$ are required. Let $\varphi_{a,i,s}$ be the a th element of $\text{vec}(\varphi_{i,s})$ and the k th moment $\mathbf{E}(\varphi_{a,i,s}^k) = \sigma_{k,a,s}$.

Assumption 3.2. (Random Coefficients)

- (a) $\varphi_{i,s}$ is a random matrix across i and over s , and IID across i for all s ,
- (b) $\varphi_{i,s}$ and $\eta_{j,t}$ are independent for all i, j, t , and s ,
- (c) $\mathbf{E}\|\varphi_{i,s}\|^4 < \infty$ for all s .
- (d) $\sum_{s=0}^{\infty} s^2 \sigma_{2,a,s} < \infty$,
- (e) $\sum_{s=0}^{\infty} s^4 (\sigma_{4,a,s})^{1/4} < \infty$,
- (f) $\varphi_i(1)\varphi_i(1)'$ is almost surely positive definite,

Conditions (a)-(e) are used by Phillips and Moon (1999b) for studying linear nonstationary panel models. Moreover, according to Lemma 1 of Phillips and Moon (1999a), Conditions (d) and (e) are sufficient for $\varphi_i(1) = \sum_{s=0}^{\infty} \varphi_{i,s} < \infty$ a.s.. Condition (f) makes sure the covariance matrix of $x_{i,t}$ and $z_{i,t}$ are well defined. Moreover, Condition (f) ensures that $x_{i,t}$ and $z_{i,t}$ are random walk processes. According to Lemma 2 from Phillips and Moon (1999b), under Assumptions 3.1(3.1') and 3.2, the process $V_{i,t}$ defined in (3.3) admits the following Beveridge-Nelson (BN) decomposition,

$$V_{i,t} = \varphi_i(1)\eta_{i,t} + \tilde{V}_{i,t-1} - \tilde{V}_{i,t} \text{ a.s.},$$

where $\tilde{V}_{i,t}$ are well defined square integrable random vectors. Hence, the partial sums of $V_{i,t}$ can be written as

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} V_{i,t} = \varphi_i(1) \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \eta_{i,t} + \frac{1}{\sqrt{T}} \tilde{V}_{i,0} - \frac{1}{\sqrt{T}} \tilde{V}_{i,[Tr]} \text{ a.s.}, \quad (3.6)$$

where $[Tr]$ denotes the integer part of Tr . Furthermore, we define

$$B_{u,i,T}^0(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_{i,t} \quad \text{and} \quad B_{v,i,T}^0(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} V_{i,t}.$$

According to Phillips and Solo (1992) and the panel functional central limit theorem (CLT) asymptotics from Phillips and Moon (1999b), the following holds for the vector Brownian motion process $(B_{u,i}, B'_{v,i})'$,

$$(B_{u,i,T}^0(r), B_{v,i,T}^{0r}(r))' \xrightarrow{d} (B_{u,i}, B'_{v,i})',$$

with both $(B_{u,i,T}^0(r), B_{v,i,T}^{0r}(r))'$ and $(B_{u,i}, B'_{v,i})'$ being defined in $D[0, 1]^3$, where $D[0, 1]$ is the space of cadlag functions on $[0, 1]$. As Park and Phillips (2001) pointed out, the above convergence in distribution can be interpreted as weak convergence in $D[0, 1]^3$ with the supremum norm topologized with the uniform topology. Then, according to the Skorodhod representation theorem, there is a common probability space¹ supporting $(B_{u,i,T}, B'_{v,i,T})'$ and $(B_{u,i}, B'_{v,i})'$ such that,

$$(B_{u,i,T}^0, B_{v,i,T}^{0r})' =^d (B_{u,i,T}, B'_{v,i,T})' \quad \text{and} \quad (B_{u,i,T}, B'_{v,i,T})' \xrightarrow{a.s.} (B_{u,i}, B'_{v,i})',$$

in $D[0, 1]^3$ with the uniform topology and $B_{v,i} = (B_{x,i}, B'_{z,i})'$. The following proofs of weak consistency are all based on $(B_{u,i,T}, B'_{v,i,T})'$. Moreover, the covariance matrix of $(B_{u,i}, B'_{v,i})'$ is written as

$$\Omega_i = \begin{bmatrix} \omega_{uu,i}^2 & \omega_{uv,i} \\ \omega_{vu,i} & \Omega_{vv,i} \end{bmatrix}.$$

According to Assumption 3.1(3.1'), $\omega_{uu,i}^2 = \sigma_{uu,i}^2$ and furthermore,

$$\omega_{vu,i} = \varphi_i(1) \begin{bmatrix} \sigma_{\varepsilon_x u,i} \\ \sigma_{\varepsilon_z u,i} \end{bmatrix} \quad \text{and} \quad \Omega_{vv,i} = \varphi_i(1) \varphi_i(1)',$$

because

$$B_{v,i} = \varphi_i(1) W_{v,i}(r).$$

Under Assumptions 3.1(3.1') and 3.2, Phillips and Moon (1999b) show that

$$\Omega = \begin{bmatrix} \omega_{uu}^2 & \omega_{uv} \\ \omega_{vu} & \Omega_{vv} \end{bmatrix} = \mathbf{E}(\Omega_i),$$

¹The original reasoning is based on assuming deterministic coefficients of the underlying moving average processes. Here, the random coefficients $\varphi_{i,t}$ are introduced with a probability space independent of above mentioned common space. Hence, the common space can be enlarged by the product measure given the independence across i .

exists with $\mathbf{E}\|\Omega_i\Omega_i'\| < \infty$. For the simplicity of notations, let

$$g(z_{i,t}, x_{i,t}, \theta) = f(z_{i,t}, \theta)x_{i,t}. \quad (3.7)$$

This chapter concentrates on the Pooled NLS estimator of Model (3.1) with the following objective function $Q_{n,T}(\theta)$,

$$Q_{n,T}(\theta) = \sum_{i=1}^n \sum_{t=1}^T (y_{i,t} - g(z_{i,t}, x_{i,t}, \theta))^2. \quad (3.8)$$

Then, the Pooled NLS estimator $\hat{\theta}_{n,T}$ is defined as the minimizer of $Q_{n,T}(\theta)$ over $\theta \in \Theta$,

$$\hat{\theta}_{n,T} = \arg \min_{\theta \in \Theta} Q_{n,T}(\theta). \quad (3.9)$$

Furthermore, let Θ denote the parameter space of θ such that θ_0 is the true value of the parameter. The following assumption is conventional in the nonlinear econometric literature to ensure the uniqueness and identifiability of the parameters (See, for example, Park and Phillips (2001)).

Assumption 3.3. (Parameter Space)

- (a) Θ is compact and convex,
- (b) θ_0 is an interior point of Θ .

Remark 3.1.

- (1) To ensure the existence of the Pooled NLS estimator, the objective function is required to be continuous on a compact and convex parameter space including θ_0 . The continuity condition is generally implied by the differentiability of the objective function which is required for the derivation of asymptotic normality. Hence, only compactness and convexity of the parameter space is assumed in Condition (a). As it will be shown later, one can derive a compact and convex set in the Euclidean space under certain restrictions on the function $g(\cdot, \theta)$.
- (2) The classical proof of consistency in nonlinear econometrics includes two basic elements: uniform convergence of the objective function on the parameter space and identifiable uniqueness of the true parameter. The uniform convergence can be derived based on existing uniform laws of large numbers of nonlinear transformations of nonstationary time series (see Park and

Phillips (1999) and Park and Phillips (2001)). As Pötscher and Prucha (1997) pointed out, the identifiable uniqueness condition is stronger than the uniqueness of θ_0 . In the absence of nuisance parameters, lower semicontinuity (hence, continuity suffices), uniqueness of θ_0 and a compact parameter space are sufficient for the identifiable uniqueness of θ_0 .

If the function f is H -regular, \dot{h} , \ddot{h} and \ddot{h} are, respectively, denoted as the limiting homogeneous functions of \dot{f} , \ddot{f} and \ddot{f} . Moreover, $\dot{\nu}$, $\ddot{\nu}$ and $\ddot{\nu}$ denote the corresponding asymptotic orders. Moreover, $\dot{\nu}$, $\ddot{\nu}$ and $\ddot{\nu}$ are only functions of \sqrt{T} because the random walk process is usually scaled by \sqrt{T} for deriving its asymptotic properties. With θ possibly being a vector, the derivative functions are also possibly vector functions. Hence, it is convenient to define $\dot{\nu}$, $\ddot{\nu}$ and $\ddot{\nu}$ as diagonal matrices with the asymptotic orders being the nonzero diagonal elements. In order to derive the asymptotic consistency property of the Pooled NSL estimator, we further require the following assumptions as in Chang et al. (2001).

Assumption I. *The function $f \in \mathcal{I}$, satisfies*

- (a) \dot{f} , \ddot{f} and \ddot{f} are I -regular,
- (b) $\int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds > 0$.

Here, we take Definition 3.3 from Park and Phillips (2001). It is obvious from the original definition that there is a trade off between smoothness and moments condition p .

Assumption H. *The function $f \in \mathcal{H}_0$, satisfies*

- (a) \dot{f} , \ddot{f} and \ddot{f} are H -regular with $|(\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu}| < \infty$,
- (b) $\int_{|s| \leq \delta} \dot{h}(s, \theta_0) \dot{h}(s, \theta_0)' ds > 0$ for all $\delta > 0$,

Remark 3.2.

- (1) \mathcal{I} and \mathcal{H}_0 refer to the class of I -regular functions and the class of H -regular functions as defined in Park and Phillips (2001), respectively. To ensure \mathcal{I} and \mathcal{H}_0 are disjoint, $\liminf_{\lambda \rightarrow \infty} \nu(\lambda) > 0$ and $h \neq 0$.
- (2) In both Assumption I and H, condition (a) gives the regularity conditions on the nonlinear regression function, while condition (b) ensures that θ_0 can be uniquely identified by the asymptotic limit of the objective function. In Section 3.3, we assume either $f \in \mathcal{I}$ or $f \in \mathcal{H}_0$ to

satisfy the conditions in Assumption I or H. Moreover, in Section 3.4, we generalize $f(\cdot, \theta)$ to allow for additive components belonging to different function classes. In fact, the specification can be generalized to a vector form, i.e., f is a vector of I - or H -regular functions. Then, the specification (3.1) is a sum of products of different functions of different nonstationary time series.

3.3 Simple Strong Co-summable Regressions

The simple co-summability relationship regarding to the model specification (3.1) is defined as the following:

Definition 3.1. (Simple Strong Co-summability) *The regression model (3.1) defines a simple co-summable relation if $y_{i,t}$ and $f(z_{i,t}; \theta)x_{i,t}$ have the same degree of summability with respect to any element of θ .*

Remark 3.3.

- (1) The definition simply implies that if $g(\cdot; \theta)$ consists of more than one additive component, each component must have the same degree of summability. In terms of co-summability, it implies that there is only one order of co-summability, and all regression components on the right hand side of Model (3.1) are $S(d)$ with $d \geq 0$. Examples of linear regressions include cointegration or fractional cointegration of regressors with the same order of integration and standard regressions with only $I(0)$ processes.
- (2) It is worth noticing that the definition is defined with respect to the parameters but not the variables of the function. For example,

$$y_{i,t} = [\alpha z_{i,t} + q(z_{i,t}; \alpha, \beta)]x_{i,t} + u_{i,t},$$

with $q(z_{i,t}; \alpha, \beta)$ being an I -regular function. According to the specification (3.1), this implies that $f(z_{i,t}; \theta) = \alpha z_{i,t} + q(z_{i,t}; \alpha, \beta)$ with $\theta = (\alpha, \beta)'$. With respect to α , the function f is actually H -regular. However, with respect to β , the function f is I -regular. Hence, this model does not satisfy the definition of simple strong co-summability.

One popular example that satisfies Definition 3.1 is the Threshold cointegration model

$$y_{i,t} = [\alpha + \beta 1\{z_{i,t} > 0\}]x_{i,t} + u_{i,t},$$

where α and $\beta 1\{z_{i,t} > 0\}$ are both H -regular functions with zero degree of homogeneity. Hence, in terms of the defined I -regular and H -regular function classes, Definition 3.1 requires that additive components of $f(\cdot; \theta)$ are all I -regular or all H -regular with the same degree of homogeneity. Hence, the results of this section do not apply to the following case,

$$y_{i,t} = [\alpha z_{i,t}^2 + \beta z_{i,t}]x_{i,t} + u_{i,t},$$

which is considered in Section 3.4. Moreover, Definition 3.1 allows different additive components of f to have parameters in common.

3.3.1 Preliminary Results

For the derivation of the results for the specification (3.1) in Section 3.2, we first need asymptotic results for sample covariances such as $\sum_{t=1}^T f(z_{i,t}, \theta)x_{i,t}$ and $\sum_{t=1}^T f(z_{i,t}, \theta)x_{i,t}u_{i,t}$ where $f(\cdot, \theta)$ is an I -regular function. To derive the asymptotic distributions, we need the following lemma.

Lemma 3.1. *Suppose that $x_{i,t}$, $z_{i,t}$ and $u_{i,t}$ satisfy Assumptions 3.1(3.1') and 3.2 with $z_{i,t}$ and $x_{i,t}$ being integrated time series defined as in (3.2) and (3.3). Let $f(z_{i,t}, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be an I -regular function. Then, the following holds, as $T \rightarrow \infty$:*

$$\frac{1}{(\sqrt{T})^{(1+k)}} \sum_{t=1}^T f(z_{i,t}, \theta)(x_{i,t})^k \xrightarrow{d} \int_{-\infty}^{\infty} f(s, \theta) ds \int_0^1 B_{x,i}^k(r) dL_{z,i}(r, 0),$$

uniformly on Θ . Moreover, if $f(s, \theta)$ is I -regular,

$$\frac{1}{(\sqrt{T})^{(1/2+k)}} \sum_{t=1}^T f(z_{i,t}, \theta)(x_{i,t})^k u_{i,t} \xrightarrow{d} \left(\int_{-\infty}^{\infty} f(s, \theta)^2 ds \int_0^1 B_{x,i}^{2k}(r) dL_{z,i}(r, 0) \right)^{1/2} W_i(1),$$

with $k > 0$, as $T \rightarrow \infty$.

$B_{x,i}$ is the Brownian motion process to which $x_{i,t}$ converges in distribution, and $L_{z,i}(r, 0)$ is the local time of $B_{z,i}$ at point 0 before time r . Furthermore, the Brownian motion W_i is independent of $B_{x,i}$ and $B_{z,i}$, and therefore of $L_{z,i}(r, 0)$. It is assumed that W_i is defined in a common probability space supporting $B_{u,i}$ and $B_{v,i}$. The following lemma ensures the existence of the Pooled LS estimator for a panel co-summable model.

Lemma 3.2. *Suppose that $x_{i,t}$, $z_{i,t}$ and $u_{i,t}$ satisfy Assumptions 3.1(3.1') and 3.2 with $z_{i,t}$ and $x_{i,t}$ are integrated time series defined as in (3.2) and (3.3). In addition, the function $g(x_{i,t}, z_{i,t}, \theta)$*

satisfies Assumption 3.3. Then, there exists a measurable function $\hat{\theta}_{n,T}$ from Y into Θ such that, for all $y = \{\{y_{i,t}\}_{t=1}^T\}_{i=1}^n$ in Y ,

$$Q_{n,T}(\hat{\theta}_{n,T}(y), y) = \inf_{\theta} Q_{n,T}(\theta, y),$$

where $Y = \prod_{i=1}^n Y_i$ with Y_i being the measurable space that contains $\{y_{i,t}\}_{t=1}^T$.

Remark 3.4.

- (1) The proof of this lemma is based on Lemma 2 from Jennrich (1969). Three sufficient conditions are compactness of Θ , continuity of $Q_{n,T}(\theta, y)$ on Θ for given y and measurability of Y . The first two conditions follow from Assumption 3.3. The construction of the measurable space Y follows from the definition of the product measure. For any i , there is a measure μ_i defined for the stochastic process $\{y_{i,t}\}_{t=1}^T$. Given $\{y_{i,t}\}_{t=1}^T$ is independent across i , the space defined by the Cartesian product $Y = \prod_{i=1}^n Y_i$ has the measure $\mu = \prod_{i=1}^n \mu_i$. Additionally, μ_i is a probability measure, hence, totally finite and μ is unique.

3.3.2 Consistency

Assumption 3.3 ensures the existence and uniqueness of $\hat{\theta}_{n,T}$. For the consistency proof, we consider the following objective function

$$R_{n,T}^*(\theta, \theta_0) = Q_{n,T}(\theta) - Q_{n,T}(\theta_0). \quad (3.10)$$

The new objective function $R_{n,T}^*$ would yield the same Pooled NLS estimates $\hat{\theta}_{n,T}$ as $Q_{n,T}$, because only a constant $Q_{n,T}(\theta_0)$ is added conditional on the data. Given the DGP adopted in this chapter, this constant term asymptotically disappears once the original objective function (3.8) is properly scaled. In addition, the asymptotic normality analysis only depends on the derivatives of the objective function. Hence, (3.8) makes no difference from (3.10) for the asymptotic normality analysis by subtracting or adding a constant term with respect to the parameters θ . Furthermore, we define

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0))^2,$$

and,

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0))u_{i,t}.$$

Then, the following lemma shows the sufficient conditions for the consistency of the Pooled NLS estimator when $T \rightarrow \infty$ followed by $n \rightarrow \infty$.

Lemma 3.3. *Suppose that $M_{n,T}$ is a strict positive function of n and T such that*

- (a) $M_{n,T}^{-1} R_{n,T,1}^*(\theta, \theta_0) \xrightarrow{p(a.s.)} R^*(\theta, \theta_0)$, and $M_{n,T}^{-1} R_{n,T,2}^*(\theta, \theta_0) \xrightarrow{p(a.s.)} 0$ uniformly in Θ as $T \rightarrow \infty$ followed by $n \rightarrow \infty$,
- (b) $R^*(\cdot, \theta_0) \geq 0$ is deterministic, continuous and has unique minimum θ_0 ,
- (c) Θ is compact.

Then,

$$\hat{\theta}_{n,T} \xrightarrow{p(a.s.)} \theta_0.$$

As the above discussion shows, the compactness of the parameter space is critical for the existence, uniqueness and consistency of the estimator. We try to use the following assumptions on the function $g(x_{i,t}, z_{i,t}; \theta)$ to construct a compact set of the parameter space, which is sufficient for the analysis.

Assumption 3.4. *Assume*

- (a) Θ is a subset of an Euclidean space and equipped with Euclidean norm,
- (b) $E\left(F\left[\inf_{\Theta} \left\|\frac{\partial g(\cdot, \theta)}{\partial \theta}\right\|^2\right]\right) \neq 0$, where

$$F\left(\inf_{\Theta} \left\|\frac{\partial g(\cdot, \theta)}{\partial \theta}\right\|^2\right) = \lim_{M_T \rightarrow \infty} \inf \sum_{t=1}^T \inf_{\Theta} \left\|\frac{\partial g(\cdot, \theta)}{\partial \theta}\right\|^2.$$

- (c) $M_{n,T} = nM_T$ with M_T is a positive function of T and $M_T \geq T$,
- (d) $Q_{n,T}(\theta)$ is almost surely strictly convex on Θ ,
- (e) $g(\cdot, \theta)$ is twice continuously differentiable on Θ .

Remark 3.5.

- (1) Sufficient conditions for Condition (b) require $g(\cdot, \theta)$ to be almost surely strict monotone on the parameter space and with finite first order partial derivative. Then, a compact set Θ exists such that $\hat{\theta}_{n,T} \in \Theta$ and $\theta_0 \in \Theta$ for all n and T .

Lemma 3.4. *Assumption 3.4 implies Assumption 3.3.*

Remark 3.6.

- (1) The above lemma indicates that the parameter space is not necessarily assumed to be compact, and if we require a bit more conditions on the nonlinear functions, it is always possible to construct a compact and convex subspace of an Euclidean space Θ such that $\hat{\theta}_{n,T} \in \Theta$ for all n and T with $\Theta = \{\theta : \|\theta - \theta_0\| \leq M\}$, where M is a strict positive number.
- (2) $\hat{\theta}_{n,T}$ exists and is unique for all n and T . The existence is proved by Lemma 3.2. The uniqueness follows from the assumption of the convexity of the objective function.

Using these preliminary results, we may now using the set of assumptions given in Assumptions 3.1 and 3.2 and 3.4 to check Conditions (a), (b) and (c) in Lemma 3.3, and state the following consistency results.

Theorem 3.1. *Let Assumptions 3.1 and 3.2 and 3.4 hold, and let $f(z_{i,t}, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ satisfy Definition 3.1 and Assumption I. Moreover, if $\int_{-\infty}^{\infty} [f(s, \theta) - f(s, \theta_0)]^2 ds > 0$ for all $\theta \neq \theta_0$, we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$R^*(\theta, \theta_0) = \left(\int_{-\infty}^{\infty} [f(s, \theta) - f(s, \theta_0)]^2 ds \right) \mathbf{E} \left(\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right),$$

with $M_{n,T} = nT^{3/2}$. Moreover,

$$\hat{\theta}_{n,T} \xrightarrow{p} \theta_0.$$

According to the expression of $R^*(\theta, \theta_0)$, Condition (b) of Lemma 3.3 can be verified by the inequality given in the theorem and Assumption 3.2 (h). $\mathbf{E} \left(\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right)$ can only be zero if $B_{x,i} = aB_{z,i}$ where a is a constant, which is ruled out by assuming almost surely nonsingular $\varphi_i(1)$ in Assumption 3.2.

Theorem 3.2. *Let Assumptions 3.1 and 3.2 and 3.4 hold, and let $f(z_{i,t}, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ satisfy Definition 3.1 and Assumption H. Moreover, if*

- (a) $\nu(\sqrt{T})$ is bounded away from zero as $T \rightarrow \infty$, and
- (b) $\int_{|s| \leq \delta} [h(s, \theta) - h(s, \theta_0)]^2 ds > 0$ for all $\theta \neq \theta_0$ and $\delta > 0$,

we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\begin{aligned} R^*(\theta, \theta_0) &= \mathbf{E} \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 B_{x,i}^2(r) dr \right) \\ &\geq \mathbf{E} \left(\sqrt{\int_{-\infty}^{\infty} [h(s, \theta) - h(s, \theta_0)]^2 L_{z,i}(1, s) ds} \left[\int_0^1 B_{x,i}^2(r) dr \right]^{-1} \right), \end{aligned}$$

with $M_{n,T} = n(T\nu(\sqrt{T}))^2$. Moreover,

$$\hat{\theta}_{n,T} \xrightarrow{P} \theta_0.$$

The nonnegative lower bound $\mathbf{E} \left(\sqrt{\int_{-\infty}^{\infty} [h(s, \theta) - h(s, \theta_0)]^2 L_{z,i}(1, s) ds} \left[\int_0^1 B_{x,i}^2(r) dr \right]^{-1} \right)$ is obtained by the Reverse Hölder inequality. According to Park and Phillips (2001), under Condition (b) in Theorem 3.2, this nonnegative lower bound is equal to zero only if $\theta = \theta_0$. Hence, Condition (b) of Lemma 3.3 can be verified.

3.3.3 Asymptotic Normality

Following the notation of (3.7), we define

$$\dot{g} = \left(\frac{\partial g}{\partial \theta_j} \right), \quad \ddot{g} = \left(\frac{\partial^2 g}{\partial \theta_j \partial \theta_l} \right), \quad \dddot{g} = \left(\frac{\partial^3 g}{\partial \theta_j \partial \theta_l \partial \theta_m} \right),$$

to be all vectors, arranged by the lexicographic ordering of their indices j , l and m which stand for the indices of the elements in θ . Sometimes, it is more convenient to define the second derivatives in matrix form as $\ddot{G} = \frac{\partial^2 g}{\partial \theta \partial \theta'}$. Clearly, \ddot{g} can be obtained by stacking the rows of \ddot{G} into a column vector. As mentioned early in Chapter 2, the asymptotic normality analysis is based on the derivatives of the objective function. Hence, according to the convention in the nonlinear econometric literature, we use $Q_{n,T}(\theta)$ instead of $R_{n,T}^*(\theta, \theta_0)$. In fact, the difference is $Q_{n,T}(\theta_0)$ which is a constant with respect to θ . Therefore, let $\dot{Q}_{n,T}$ and $\ddot{Q}_{n,T}$ be the first and second derivatives of $Q_{n,T}$ with respect to θ defined as $\dot{Q}_{n,T} = \partial Q_{n,T} / \partial \theta$ and $\ddot{Q}_{n,T} = \partial^2 Q_{n,T} / \partial \theta \partial \theta'$, then

$$\begin{aligned} \dot{Q}_{n,T} &= - \sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta) (y_{i,t} - g(z_{i,t}, x_{i,t}, \theta)), \\ \ddot{Q}_{n,T} &= \sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta) \dot{g}(z_{i,t}, x_{i,t}, \theta)' - \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta) (y_{i,t} - g(z_{i,t}, x_{i,t}, \theta)), \end{aligned}$$

where an unimportant constant is ignored. The asymptotic distribution of $\hat{\theta}_{n,T}$ can be obtained by the first order Taylor expansion of $\dot{Q}_{n,T}$, which is written as

$$\dot{Q}_{n,T}(\hat{\theta}_{n,T}) = \dot{Q}_{n,T}(\theta_0) + \ddot{Q}_{n,T}(\theta_{n,T})(\hat{\theta}_{n,T} - \theta_0), \quad (3.11)$$

where $\theta_{n,T}$ lies on the line segment connecting $\hat{\theta}_{n,T}$ and θ_0 . If $\hat{\theta}_{n,T}$ is an interior solution to the minimization problem (3.9), then $\dot{Q}_{n,T}(\hat{\theta}_{n,T}) = 0$. For an appropriately chosen normalizing sequence $C_{n,T}$, which can be viewed as a scalar or diagonal matrix, we can prove that $C_{n,T}^{-1}\dot{Q}_{n,T}(\theta_0) \xrightarrow{d} \dot{Q}(\theta_0)$ for some random vector $\dot{Q}(\theta_0)$. Furthermore, let

$$\ddot{Q}_{n,T}^0(\theta_0) = \sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta_0) \dot{g}(z_{i,t}, x_{i,t}, \theta_0)',$$

then $C_{n,T}^{-1}\ddot{Q}_{n,T}^0(\theta_0)C_{n,T}^{-1'} \xrightarrow{p} \ddot{Q}(\theta_0)$ for some random matrix $\ddot{Q}(\theta_0)$. Therefore, under suitable conditions that ensure $C_{n,T}^{-1}\ddot{Q}_{n,T}(\theta_{n,T})C_{n,T}^{-1'} = C_{n,T}^{-1}\ddot{Q}_{n,T}^0(\theta_0)C_{n,T}^{-1'} + o_p(1)$ and $\ddot{Q}(\theta_0) > 0$ a.s., the following result may be expected from Eq. (3.11):

$$\begin{aligned} C_{n,T}^{-1}(\hat{\theta}_{n,T} - \theta_0) &= - \left(C_{n,T}^{-1}\ddot{Q}_{n,T}(\theta_{n,T})C_{n,T}^{-1'} \right)^{-1} C_{n,T}^{-1}\dot{Q}_{n,T}(\theta_0) \\ &= \left(C_{n,T}^{-1}\ddot{Q}_{n,T}^0(\theta_0)C_{n,T}^{-1'} \right)^{-1} C_{n,T}^{-1}\dot{Q}_{n,T}(\theta_0) + o_p(1) \\ &\xrightarrow{d} - (\ddot{Q}(\theta_0))^{-1}\dot{Q}(\theta_0), \end{aligned} \quad (3.12)$$

as $T \rightarrow \infty$ followed by $n \rightarrow \infty$. Then, define

$$J_{n,T} = C_{n,T}^{-1}\ddot{Q}_{n,T}^0(\theta_0)C_{n,T}^{-1},$$

and

$$D_{n,T} = C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta_0) u_{i,t} C_{n,T}^{-1},$$

so that $J_{n,T} = C_{n,T}^{-1}\ddot{Q}_{n,T}(\theta_0)C_{n,T}^{-1} + D_{n,T}$, and let $L_{n,T} = -C_{n,T}^{-1}\dot{Q}_{n,T}(\theta_0)$. Given the consistency results in Theorem 3.1 and Theorem 3.2, and the six sufficient conditions in Page 138 of Park and Phillips (2001) for asymptotic normality can be concluded as the following lemma.

Lemma 3.5. *Suppose that, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$,*

$$(a) \ (J_{n,T}, L_{n,T}) \xrightarrow{d} (J, L),$$

$$(b) \ D_{n,T} = o_p(1),$$

$$(c) \ J > 0 \text{ a.s.},$$

$$(d) \ \dot{Q}_{n,T}(\hat{\theta}_{n,T}) = 0 \text{ a.s.},$$

$$(e) \ C_{n,T}^{-1}(\ddot{Q}_{n,T}(\theta_{n,T}) - \ddot{Q}_{n,T}(\theta_0))C_{n,T}^{-1} \xrightarrow{P} 0 \text{ uniformly in } \theta \in \Theta_T.$$

Then $C_{n,T}(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{d} J^{-1}L$.

Recall that $\dot{\nu}$ and \dot{h} are, respectively, the *asymptotic order* and the *limiting homogeneous function* of \dot{f} that is the first order derivatives of f with respect to θ . The asymptotic distributions are given by the following theorems.

Theorem 3.3. *Let Assumptions 3.1, 3.2 and 3.4 hold, and let $f(z_{i,t}, \theta)$ satisfy the conditions in Theorem 3.1. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\sqrt{n}^{3/4} \sqrt{T}(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{d} \mathbf{N}(0, \Sigma^{-1}),$$

where

$$\Sigma = \int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right].$$

Theorem 3.4. *Let Assumptions 3.1', 3.2 and 3.4 hold, and let $f(z_{i,t}, \theta)$ satisfy the conditions in Theorem 3.2. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\sqrt{n}T\dot{\nu}(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{d} \mathbf{N}(0, \Sigma_H^{-1}),$$

where

$$\Sigma_H = \mathbf{E} \left[\int_0^1 \dot{h}(B_{z,i}, \theta_0) \dot{h}(B_{z,i}, \theta_0)' B_{x,i}^2 dr \right].$$

To prove Theorems 3.3 and 3.4, it is sufficient to show that all conditions in Lemma 3.4 are satisfied. Conditions (a) and (b) follow directly from Lemma 5 from Chang et al. (2001) or Lemma

3.1 given above. Condition (c) is satisfied by the identification condition in Assumption I and H. Only condition (e) needs to be verified.

For the I -regular case, the same assumptions are sufficient for both consistency and asymptotic normality of the Pooled LS estimator. In particular, $\sqrt{n}^{3/4}\sqrt{T}$ can be obtained even if the regressors are weakly exogenous. However, if the function is H -regular and the regressors are weakly exogenous, we can only obtain consistency but not necessary $\sqrt{nT}\dot{\nu}$ consistency. As Theorem 3.4 shows $\sqrt{nT}\dot{\nu}$ consistency is obtained under Assumption 3.1' with strictly exogenous regressors. This result is in line with that obtained for the linear panel cointegration models in which the linear function is a special case of H -regular functions.

3.3.4 Estimation of the Covariance Matrix

In this section, we consider the following estimator for the covariance matrices Σ and Σ_H and prove its weak consistency,

$$\hat{\Sigma}_{n,T} = C_{n,T}^{-1} \begin{bmatrix} \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\beta}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{c}_{n,t}} \\ \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\beta}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{c}_{n,t}} \\ \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} \\ \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{c}_{n,t}^2} \end{bmatrix} C_{n,T}^{-1}, \quad (3.13)$$

where $\hat{Q}_{n,T}$ is the objective function being evaluated at the estimates. Then, the following corollaries show the consistency of $\hat{\Sigma}_{n,T}$.

Corollary 3.1. *Let the assumptions of Theorem 3.3 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma,$$

where $C_{n,T} = \sqrt{n}^{3/4}\sqrt{T}$ and

$$\Sigma = \int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right].$$

Corollary 3.2. *Let the assumptions of Theorem 3.4 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma_H,$$

where $C_{n,T} = \sqrt{nT}\dot{\nu}$

$$\Sigma_H = \mathbf{E} \left[\int_0^1 \dot{h}(B_{z,i}, \theta_0) \dot{h}(B_{z,i}, \theta_0)' B_{x,i}^2 dr \right].$$

3.4 Multiple Strong Co-summable Regressions

In this section, we define multiple strong co-summability regressions as

Definition 3.2. (Multiple Strong Co-summability) *The regression model (3.1) defines a multiple co-summable relation if $f(z_{i,t}; \theta)$ is of an additive form and has different degrees of summability with respect to different elements of θ , and the degree of summability of $y_{i,t}$ is equal to the highest degree of summability of $f(z_{i,t}; \theta)x_{i,t}$ with respect to the elements of θ .*

Remark 3.7.

- (1) The definition simply implies that if $f(\cdot; \theta)$ consists of more than one additive component, different components may have different degrees of summability. In terms of co-summability, it says that $y_{i,t}$ and at least one component of $f(\cdot; \theta)$ are co-summable with degree d_1 which yields a residual term $S(d_2)$ and $d_1 > d_2 \geq 0$. Then, this residual term is co-summable with other $S(d_2)$ terms to yield an $I(0)$ residual term. We can iterate this argument for any finite K steps which defines a K – stage co-summable regression. One typical corresponding case in linear regression consists of two cointegrated $I(2)$ processes of which the residual term is cointegrated with another $I(1)$ process.
- (2) The example given in Item (2) of Remark 3.1 actually defines a 2 – stage co-summable relation. Hence, $y_{i,t}$ and the first component on the right hand side are $S(3/2)$. The second component is $S(1/2)$. Accordingly, we expect that $f(\cdot; \theta)$ consists of at least two additive components. Otherwise, Model (3.1) only defines a simple co-summable relationship.

According to Definition 3.2, we further specify Model (3.1) as follows:

$$y_{i,t} = \{f_1(z_{i,t}; \theta_1, \theta_2, \theta_3) + f_2(z_{i,t}; \theta_2, \theta_3, \theta_4)\}x_{i,t} + u_{i,t}, \quad (3.14)$$

with $\theta = (\theta'_1, \theta'_2, \theta'_3, \theta'_4)'$ and $f(z_{i,t}; \theta) = f_1(z_{i,t}; \theta_1, \theta_2, \theta_3) + f_2(z_{i,t}; \theta_2, \theta_3, \theta_4)$. Simple Strong Co-summable Regression implies that f_1 and f_2 have the same degree of summability with respect to

any parameter of θ_1 , θ_2 , θ_3 and θ_4 . Moreover, we distinguish the following two cases.

Assumption HH:

- (a) Assume that $f_1(\cdot; \theta_1, \theta_2, \theta_3)$ satisfies Assumption H with respect to $(\theta'_1, \theta'_2)'$, and the degree of homogeneity is ν_1 ;
- (b) Assume that $f_1(\cdot; \theta_1, \theta_2, \theta_3)$ satisfies Assumption H with respect to θ_3 , and the degree of homogeneity is ν_3 ;
- (c) Assume that $f_2(\cdot; \theta_2, \theta_3, \theta_4)$ satisfies Assumption H with respect to $(\theta_2, \theta'_3, \theta'_4)'$, and the degree of homogeneity is ν_2 ;
- (d) $\frac{\nu_2}{\nu_1} = o_p(1)$.

An example satisfying Assumption HH is $y_{i,t} = \{\alpha z_{i,t}^2 + \beta z_{i,t}\}x_{i,t} + u_{i,t}$. Regarding the parametrization with respect to θ_3 , we have the following three different cases:

- Case 1: If f_1 is an H – regular function with respect to θ_3 with degree of homogeneity $\nu_1 > \nu_3 > \nu_2$. The model should be reparametrized such that f_2 is the remainder function of f_1 with respect to θ_3 . Namely, the model can be reparametrized as the sum of three functions f_1 , f_2 and f_3 with f_3 being an H – regular with respect to θ_3 with the degree of homogeneity being ν_3 . The arguments of Theorem 5.1 can be iteratively extended to give similar results in this case. The only example of such a case is that θ_3 appears both in the remainder term of f_1 and f_2 . Otherwise, f_1 would be an H – regular function with respect to θ_3 of ν_1 degree of homogeneity.
- Case 2: If f_1 is an H – regular function with respect to θ_3 with degree of homogeneity $\nu_3 < \nu_2$. Then the appearance of θ_3 in f_1 has no influence on the consistent estimation and identification determined by f_2 ;
- Case 3: If f_1 is an H – regular function with respect to θ_3 with degree of homogeneity $\nu_3 = \nu_2$. The consistency of estimation and identification of θ_3 are determined by both the remainder term of f_1 and f_2 .

Next, we consider the multiple strong co-summable regression consisting of both H – regular and I – regular functions.

Assumption HI:

- (a) Assume that $f_1(\cdot; \theta_1, \theta_2, \theta_3)$ satisfies Assumption H with respect to $(\theta'_1, \theta'_2)'$, and the degree of homogeneity is ν_1 ;
- (b) Assume that $f_2(\cdot; \theta_2, \theta_3, \theta_4)$ satisfies Assumption I with respect to $(\theta_2, \theta'_3, \theta'_4)'$;
- (c) The remainder term of f_1 satisfies Assumption I with respect to θ_3 .

The zero function satisfies Assumption I in Section 3.2. Hence, if θ_3 does not appear in the remainder term, namely $f_1(\cdot; \theta_1, \theta_2)$, condition [(c)] is satisfied. An example satisfying Assumption HI is the Smooth Transition Regression model with the logistic transition function.

3.4.1 Consistency

Recall that h generally denotes the *limiting homogeneous function* of a H – *regular* function. For simplicity of the notation, we define $h_{r_1}^0 = h_{r_1}(s; \theta_{1,0}, \theta_{2,0}, \theta_{3,0})$, $\bar{h}_{r_1} = h_{r_1}(s; \theta_{1,0}, \theta_{2,0}, \theta_3)$, $h_2^0 = h_2(s; \theta_{2,0}, \theta_{3,0}, \theta_{4,0})$, $\bar{h}_2 = h_2(s; \theta_{2,0}, \theta_3, \theta_4)$, $\bar{h}_{r_1}^B = h_{r_1}(B_{z,i}; \theta_{1,0}, \theta_{2,0}, \theta_3)$, $h_{r_1}^{B,0} = h_{r_1}(B_{z,i}; \theta_{1,0}, \theta_{2,0}, \theta_{3,0})$, $\bar{h}_2 = h_2(B_{z,i}; \theta_{2,0}, \theta_3, \theta_4)$ and $h_2^{B,0} = h_2(B_{z,i}; \theta_{2,0}, \theta_{3,0}, \theta_{4,0})$, where h_{r_1} is the homogeneous function of the remainder term of f_1 with respect to θ_3 . $h_{r_1}^0$, \bar{h}_{r_1} , h_2^0 , \bar{h}_2 are functions with the argument s for all $s \in \mathbb{R}$. Without loss of generality, we only need to consider cases 2 and 3 in this chapter. In fact, case 3 can be treated as a special case of case 2.

Theorem 3.5. *Let Assumptions 3.1, 3.2 and 3.4 hold, and let $f(z_{i,t}; \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as Model (3.14) and satisfy Assumption HH. Moreover, if*

- (a) $\nu_i(\sqrt{T})$ and $\dot{\nu}_i(\sqrt{T})$ are bounded away from zero as $T \rightarrow \infty$, for $i = 1, 2$,
- (b) $\nu_3 = \nu_2$, $\frac{\dot{\nu}_2}{\nu_1 \nu_2 T}$ is $o_p(1)$ and $\frac{\dot{\nu}_1}{\nu_1 \nu_2 T}$ is $o_p(1)$,
- (c) $\int_{|s| \leq \delta} [h_1(s; \theta_1, \theta_2) - h_1(s; \theta_{1,0}, \theta_{2,0})]^2 ds > 0$ for all $(\theta'_1, \theta'_2)' \neq (\theta'_{1,0}, \theta'_{2,0})'$ and $\delta > 0$,
- (d) $\int_{|s| \leq \delta} [\bar{h}_{r_1} - h_{r_1}^0 + \bar{h}_2 - h_2^0]^2 ds > 0$ for all $(\theta'_3, \theta'_4)' \neq (\theta'_{3,0}, \theta'_{4,0})'$ and $\delta > 0$,

we then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\begin{aligned}
 R^*(\theta_1, \theta_2, \theta_{1,0}, \theta_{2,0}) &= \mathbf{E} \left(\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]^2 B_{x,i}^2(r) dr \right) \\
 &\geq \mathbf{E} \left(\sqrt{\int_{-\infty}^{\infty} [h_1(s; \theta_1, \theta_2) - h_1(s; \theta_{1,0}, \theta_{2,0})]^2 L_{z,i}(1, s) ds} \left[\int_0^1 B_{x,i}^2(r) dr \right]^{-1} \right),
 \end{aligned}$$

with $M_{1,n,T} = n(T\nu_1(\sqrt{T}))^2$, and

$$\begin{aligned} R^*(\theta_{1,0}, \theta_{2,0}, \theta_3, \theta_4, \theta_{3,0}, \theta_{4,0}) &\xrightarrow{P} \mathbf{E} \left(\int_0^1 [\bar{h}_{r_1}^B - h_{r_1}^{B,0} + \bar{h}_2^B - h_2^{B,0}]^2 B_{x,i}^2(r) dr \right) \\ &\geq \mathbf{E} \left(\sqrt{\int_{-\infty}^{\infty} [\bar{h}_{r_1} - h_{r_1}^0 + \bar{h}_2 - h_2^0]^2 L_{z,i}(1, s) ds} \left[\int_0^1 B_{x,i}^2(r) dr \right]^{-1} \right), \end{aligned}$$

with $M_{2,n,T} = n(T\nu_2(\sqrt{T}))^2$. Moreover,

$$(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' \xrightarrow{P} (\theta'_{1,0}, \theta'_{2,0})',$$

at rate $(T\nu_1)^{-1}$, and

$$(\hat{\theta}'_{3,n,T}, \hat{\theta}'_{4,n,T})' \xrightarrow{P} (\theta'_{3,0}, \theta'_{4,0})',$$

at rate $(T\nu_2)^{-1}$.

Remark 3.8.

- (1) Iterating the arguments, the results can be generalized to the K -stage. If θ_3 does not appear in the remainder term of f_1 , namely in case (3), the identification condition only requires $\bar{h}_2 - h_2^0$ and only $\bar{h}_2^B - h_2^{B,0}$ appears in the limit. Note: If θ_3 appears in h_1 , f_1 is H -regular with respect to θ_3 but not $\nu_3 < \nu_1$. Moreover, the argument in Theorem 3.5 implies that the multiple strong co-summable regressions can be consistently specified in K steps. Namely, there exist consistent K -stage NLS estimators. However, the K -stage NLS estimator may have the rate of convergence being different from the normal NLS estimator, which can be shown by the Smooth Transition Co-summable regression. Namely, the K -stage NLS estimator may not be as efficient as the NLS estimator. The linear cointegration is a special case where the K -stage NLS estimator and the NLS estimator have the same rate of convergence, i.e., the Engle-Granger two-step method.
- (2) The logic of Theorem 3.5 is based on the argument that scaling the first order derivatives of the objective function by different functions of T and n would result in the same set of estimates $\hat{\theta}_{n,T}$. Suppose there are two sets of parameters θ_1 and θ_2 , and f_1 being $S(3/2)$ and f_2 being $S(0)$, we consider the following model and objective functions,

$$y_{i,t} = f_1(\theta_{1,0}, x_{i,t}) + f_2(\theta_{2,0}, z_{i,t}) + u_{i,t},$$

$$Q_{n,T}(\theta_{1,0}, \theta_{2,0}; \theta_1, \theta_2), \quad (3.15)$$

$$\frac{1}{nT^2} Q_{n,T}(\theta_{1,0}, \theta_{2,0}; \theta_1, \theta_2), \quad (3.16)$$

and

$$\frac{1}{n\sqrt{T}} Q_{n,T}(\theta_{1,0}, \theta_{2,0}; \theta_1, \theta_2). \quad (3.17)$$

They have the same set of optimal solutions for any finite T because the scalars, $\frac{1}{nT}$ and $\frac{1}{n\sqrt{T}}$, disappear in the first order conditions. Hence, if we consider the second objective function, we can have the consistency of $\hat{\theta}_{1,n,T}$. Given the values of $\hat{\theta}_{1,n,T}$, we know that

$$\frac{1}{n\sqrt{T}} Q_{n,T}(\theta_{1,0}, \theta_{2,0}; \hat{\theta}_{1,n,T}, \theta_2) \quad (3.18)$$

also has the same optimal solution for θ_2 as that of (3.15)-(3.17) if the optimal solution is unique which can be guaranteed by Assumption 3.4. Hence, the one-shot optimal solution for the objective function (3.15) can be separately treated according to (3.16) and (3.17). By (3.16), we obtain that $\hat{\theta}_{1,n,T}$ is consistent and $\theta_{2,0}$ is not identified. By (3.18), we obtain that, given the consistent estimator of $\theta_{1,0}$, $\hat{\theta}_{2,n,T}$ is consistent, where we can temporarily treat $\hat{\theta}_{1,n,T}$ as *known*. Namely, the estimate of (3.15) is equivalent to the estimates of (3.16) and (3.18).

- (2) In order to distinguish the one-shot NLS estimation with the K -stage NLS estimation, we summarize as the follows: For the one-shot NLS estimation, we estimate the complete model at once but the theoretical analysis of the consistency property is done by treating different components of the model separately. For the K -stage NLS estimation, we estimate the model separately by K steps and the theoretical analysis is also stepwise. There are similarities in the logic of the analysis of the two estimators but generically they are different.

We define $\bar{r}_1 = r_1(s; \theta_{1,0}, \theta_{2,0}, \theta_3)$, $r_1^0 = r_1(s; \theta_{1,0}, \theta_{2,0}, \theta_{3,0})$, $f_2^0 = f_2(s; \theta_{2,0}, \theta_{3,0}, \theta_{4,0})$ and $\bar{f}_2 = f_2(s; \theta_{2,0}, \theta_3, \theta_4)$, where r_1 is the remainder term of f_1 .

Theorem 3.6. *Let Assumptions 3.1, 3.2 and 3.4 hold, and let $f(z_{i,t}, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as Model (3.14) and satisfy Assumption HI. Moreover, if*

$$(a) \quad \frac{\nu_1}{\nu_1 T^{3/4}} \text{ is } o_p(1),$$

(b) $\int_{|s| \leq \delta} [h_1(s; \theta_1, \theta_2) - h_1(s; \theta_{1,0}, \theta_{2,0})]^2 ds > 0$ for all $(\theta'_1, \theta'_2)' \neq (\theta'_{1,0}, \theta'_{2,0})'$ and $\delta > 0$,

(c) $\int_{-\infty}^{\infty} [\bar{r}_1 - r_1^0 + \bar{f}_2 - f_2^0]^2 ds > 0$ for all $(\theta'_3, \theta'_4)' \neq (\theta'_{3,0}, \theta'_{4,0})'$,

we then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\begin{aligned} R^*(\theta_1, \theta_2, \theta_{1,0}, \theta_{2,0}) &= \mathbf{E} \left(\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]^2 B_{x,i}^2(r) dr \right) \\ &\geq \mathbf{E} \left(\sqrt{\int_{-\infty}^{\infty} [h_1(s; \theta_1, \theta_2) - h_1(s; \theta_{1,0}, \theta_{2,0})]^2 L_{z,i}(1, s) ds} \left[\int_0^1 B_{x,i}^2(r) dr \right]^{-1} \right), \end{aligned}$$

with $M_{1,n,T} = n(T\nu_1(\sqrt{T}))^2$, and

$$R^*(\theta_{1,0}, \theta_{2,0}, \theta_3, \theta_4, \theta_{3,0}, \theta_{4,0}) = \int_{-\infty}^{\infty} [\bar{r}_1 - r_1^0 + \bar{f}_2 - f_2^0]^2 ds^2 \mathbf{E} \left[\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right],$$

with $M_{2,n,T} = n(T)^{3/2}$. Moreover,

$$(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' \xrightarrow{P} (\theta'_{1,0}, \theta'_{2,0})',$$

at rate $(T\nu_1)^{-1}$, and

$$(\hat{\theta}'_{3,n,T}, \hat{\theta}'_{4,n,T})' \xrightarrow{P} (\theta'_{3,0}, \theta'_{4,0})',$$

at rate $(\sqrt{n}T^{3/4})^{-1}$.

The remainder term of f_1 plays an important role based on the degree of summability being equal to or smaller than that of the I - regular function.

3.4.2 Asymptotic Normality

The difference between the results of Section 3.4 and those of Section 3.5 mainly lies in the arguments of asymptotic consistency. The results in Section 3.4 show that if there are additive components of $f(z_{i,t}, \theta)$ with the same order of summability, the parameters in one component cannot be consistently estimated without specifying the others. While if the additive components have different orders of summability, the parameters in the components with higher orders of summability can be consistently estimated without specifying the lower order terms. Moreover, they can even be estimated without correctly specifying the model, which implies a K - step consistent procedure with K depending on the number of different summability orders. As we notice in further study, the K - step may

not be efficient in terms of the rate of convergence. Three examples are given here to illustrate the phenomena.

Example 3.1.

$$y_{i,t} = [\alpha z_{i,t}^2 + \beta z_{i,t}]x_{i,t} + u_{i,t},$$

where α can be consistently estimated without specifying $\beta z_{i,t}x_{i,t}$ but the estimator will not be as efficient as the one for the correctly specified model.

Example 3.2.

$$y_{i,t} = \alpha x_{i,t} + \beta f_I(z_{i,t}) + u_{i,t},$$

where f_I is an I -regular function. In this case, $f_I(z_{i,t})$ is asymptotically orthogonal to $x_{i,t}$, hence α will be estimated both consistently and efficiently without specifying the second component even if the regressors $z_{i,t}$ and $x_{i,t}$ are correlated.

Example 3.3.

$$y_{i,t} = \alpha f_I^x(x_{i,t}) + \beta f_I^z(z_{i,t}) + u_{i,t},$$

where $f_I^x(x_{i,t})$ and $f_I^z(z_{i,t})$ are both I -regular functions of different regressors. α and β will both be consistently and efficiently estimated without specifying the other.

From the above examples, it appears that this asymptotically orthogonality property actually depends on the summability order of the cross-product of different components in the regression. For simplicity, we define $\dot{f}_{i,j} = \frac{\partial f_i}{\partial \theta_j}$, for $i = 1, 2$ and $j = 1, 2, 3, 4$, and $\dot{f}_{i,j,0}$ is defined at θ_0 .

Theorem 3.7. *Let Assumptions 3.1', 3.2 and 3.4 hold, and let $f(z_{i,t}; \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as Model (3.14) and satisfy Assumption HH. Moreover, if*

(a) *all assumptions in Theorem 3.5 are satisfied,*

(b) $\frac{\dot{v}_2}{\dot{v}_1} \xrightarrow{P} 0$, $\frac{\ddot{v}_2}{\dot{v}_1} \xrightarrow{P} 0$, $\ddot{v}_3 = \ddot{v}_2$ and $\dot{v}_3 = \dot{v}_2$,

(c)

$$\int_{|s| \leq \delta} \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix} \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix}' ds > 0,$$

for all $\delta > 0$, we then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\sqrt{n}TC_T(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{d} \mathbf{N}(0, \Sigma_{HH}^{-1}),$$

where

$$C_T = \begin{bmatrix} \dot{\nu}_1 & 0 & 0 & 0 \\ 0 & \dot{\nu}_1 & 0 & 0 \\ 0 & 0 & \dot{\nu}_2 & 0 \\ 0 & 0 & 0 & \dot{\nu}_2 \end{bmatrix},$$

and

$$\Sigma_{HH} = \mathbf{E} \left\{ \int_0^1 \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix} \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix}' B_{x,i}^2 dr \right\}.$$

Apparently, Condition (b) in Theorem 3.7 is needed for the derivation because we consider a general class of functions in the theorem. Hence, it is very difficult to come up with an example that fails the condition if not impossible. However, in practice, most popular H -regular functions such as the polynomial functions satisfy this condition.

Theorem 3.8. *Let Assumptions 3.1', 3.2 and 3.4 hold, and let $f(z_{i,t}; \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as Model (3.14) and satisfy Assumption HI. Moreover, if*

(a) *all assumptions in Theorem 3.6 are satisfied,*

(b)

$$\int_{|s| \leq \delta} \begin{pmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \end{pmatrix} \begin{pmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \end{pmatrix}' ds > 0,$$

for all $\delta > 0$, and

$$\int_{-\infty}^{\infty} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix}' ds > 0,$$

we then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\sqrt{n}T\dot{\nu}_1 \begin{pmatrix} \hat{\theta}_{1,n,T} - \theta_{1,0} \\ \hat{\theta}_{2,n,T} - \theta_{2,0} \end{pmatrix} \xrightarrow{d} \mathbf{N}\left(0, \Sigma_{HI,1}^{-1}\right),$$

where

$$\Sigma_{HI,1} = \mathbf{E} \left[\int_0^1 \begin{pmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \end{pmatrix} \begin{pmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \end{pmatrix}' B_{x,i}^2 dr \right].$$

Moreover,

$$\sqrt{n}T^{3/4} \begin{pmatrix} \hat{\theta}_{3,n,T} - \theta_{3,0} \\ \hat{\theta}_{4,n,T} - \theta_{4,0} \end{pmatrix} \xrightarrow{d} \mathbf{N}\left(0, \Sigma_{HI,2}^{-1}\right),$$

where

$$\Sigma_{HI,2} = \int_{-\infty}^{\infty} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix}' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right].$$

3.4.3 Estimation of the Covariance Matrix

We define the estimator $\hat{\Sigma}_{n,T}$ for the covariance matrices Σ_{HH} and Σ_{HI} as (3.13) where.

$$\Sigma_{HI} = \begin{bmatrix} \Sigma_{HI,1} & 0 \\ 0 & \Sigma_{HI,2} \end{bmatrix}. \quad (3.19)$$

Then, the following corollaries show the consistency of $\hat{\Sigma}_{n,T}$ for multiple strong co-summable regression models.

Corollary 3.3. *Let the assumptions of Theorem 3.7 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma_{HH},$$

where $C_{n,T} = \sqrt{n}TC_T$ with C_T and Σ_{HH} being defined in Theorem 3.7.

Corollary 3.4. *Let the assumptions of Theorem 3.8 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma_{HI},$$

where $C_{n,T} = \sqrt{n}C_T$ with

$$C_T = \begin{bmatrix} T\dot{\nu}_1 & 0 & 0 & 0 \\ 0 & T\dot{\nu}_1 & 0 & 0 \\ 0 & 0 & \sqrt[3]{4T} & 0 \\ 0 & 0 & 0 & \sqrt[3]{4T} \end{bmatrix},$$

and Σ_{HI} being defined in (3.19).

3.5 Smooth Transition Co-summable Regressions

In this section, we derive the asymptotic distribution of the Pooled NLS estimator for the multiple strong co-summable regression defined by the Smooth Transition Regression models. Besides, consistent estimators for the covariance matrix are proposed for inference. The smooth transition regression model considered is as follows:

$$y_{i,t} = \{\alpha + \beta f(z_{i,t}; \gamma, c)\}x_{i,t} + u_{i,t}, \quad (3.20)$$

where $x_{i,t}$ and $z_{i,t}$ are the $S(1)$ strongly exogenous variables, and $u_{i,t}$ is a martingale difference process. Furthermore, $z_{i,t}$ is named indicator or transition variable which determines the prevailing regime at time t . The transition function $f(z_{i,t}; \gamma, c)$ in (3.20) is a continuous function that is bounded between 0 and 1. In this chapter, $f(z_{i,t}; \gamma, c)$ is chosen to be the logistic function according to Granger and Teräsvirta (1993):

$$f(z_{i,t}; \gamma, c) = \frac{1}{1 + \exp(-\gamma(z_{i,t} - c))}, \quad (3.21)$$

with $\gamma > 0$ which is an identifying restriction. The logistic function $f(z_{i,t}; \gamma, c)$ can be viewed as a cumulative distribution function. Hence, according to Table 2.2, it is H -regular with the *limit homogeneous function* $1\{z_{i,t} \geq 0\}$ and the *remainder function* $r(z_{i,t}, \gamma, c) = 1_{\mathbb{R}_-} \frac{1}{1 + \exp(-\gamma(z_{i,t} - c))} - 1_{\mathbb{R}_+} \frac{1}{1 + \exp(\gamma(z_{i,t} - c))}$ being I -regular. Accordingly, in order to derive the asymptotic property of the Pooled NLS estimator, Model (3.20) is reparametrized in an additive form as:

$$y_{i,t} = \alpha x_{i,t} + \beta x_{i,t} 1_{\mathbb{R}_+} + \beta x_{i,t} [1_{\mathbb{R}_-} \frac{1}{1 + \exp(-\gamma(z_{i,t} - c))} - 1_{\mathbb{R}_+} \frac{1}{1 + \exp(\gamma(z_{i,t} - c))}] + u_{i,t}, \quad (3.22)$$

where $1_{\mathbb{R}_+} = 1\{z_{i,t} \geq 0\}$ and $1_{\mathbb{R}_-} = 1\{z_{i,t} \leq 0\}$. The following steps establish (3.22) from (3.20):

$$y_{i,t} = \alpha x_{i,t} + \beta x_{i,t} 1_{\mathbb{R}_+} + [\beta x_{i,t} f(z_{i,t}; \gamma, c) - \beta x_{i,t} 1_{\mathbb{R}_+}] + u_{i,t},$$

$$y_{i,t} = \alpha x_{i,t} + \beta x_{i,t} 1_{\mathbb{R}_+} + \beta x_{i,t} [1_{\mathbb{R}_-} f(z_{i,t}; \gamma, c) + 1_{\mathbb{R}_+} [f(z_{i,t}; \gamma, c) - 1]] + u_{i,t}.$$

Hence, Model (3.20) is

$$y_{i,t} = f_1(x_{i,t}; \alpha) + f_2(x_{i,t}; \beta) f_3(z_{i,t}) + f_2(x_{i,t}; \beta) f_4(z_{i,t}; \gamma, c) + u_{i,t}, \quad (3.23)$$

where $f_1(x_{i,t}; \alpha) = \alpha x_{i,t}$ and $f_2(x_{i,t}; \beta) = \beta x_{i,t}$ are *homogeneous functions* of $x_{i,t}$ with order one. $f_3(z_{i,t}) = 1\{z_{i,t} \geq 0\}$ is a *homogeneous function* of $z_{i,t}$ with order zero. Moreover, $f_4(z_{i,t}; \gamma, c) = 1_{\mathbb{R}_-} \frac{1}{1 + \exp(-\gamma(z_{i,t} - c))} - 1_{\mathbb{R}_+} \frac{1}{1 + \exp(\gamma(z_{i,t} - c))}$ is an *integrable function* of $z_{i,t}$. Model (3.23) is a special case of the multiple strong co-summable regression defined by specification (3.1) in the beginning of Section 3.3.

3.5.1 Consistency and Asymptotic Normality

The proofs of the following corollaries are based on the representation (3.23) for the consistency of the Pooled NLS estimator of the parameters in Model (3.20).

Corollary 3.5. *Let Assumptions 3.1, 3.2 and 3.4 hold respectively for $x_{i,t}$, $z_{i,t}$, $u_{i,t}$ and f_k with $k = 1, 2, 3, 4$. Moreover, if*

$$\int_{|s| \leq \delta} [(\alpha_0 - \alpha) + (\beta_0 - \beta) 1\{s > 0\}]^2 ds > 0,$$

for all $\theta \neq \theta_0$ and $\delta > 0$, we then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$R^*(\theta, \theta_0) = \mathbf{E} \left[\int_0^1 [(\alpha_0 - \alpha) + (\beta_0 - \beta) 1\{B_{z,i} > 0\}]^2 B_{x,i}^2(r) dr \right],$$

with $M_{n,T} = nT^2$. Moreover, for the Pooled NLS estimators $\hat{\alpha}_{n,T}$ and $\hat{\beta}_{n,T}$:

$$\hat{\alpha}_{n,T} \xrightarrow{p} \alpha_0,$$

$$\hat{\beta}_{n,T} \xrightarrow{p} \beta_0.$$

$R^*(\theta, \theta_0)$ has a nonnegative lower bound as in Theorem 3.6. The asymptotic results in Corollary 3.5 show that the estimators $\hat{\alpha}_{n,T}$ and $\hat{\beta}_{n,T}$ actually have the same asymptotic properties as those based on,

$$y_{i,t} = \alpha x_{i,t} + \beta 1\{z_{i,t} > 0\}x_{i,t} + \tilde{u}_{i,t}, \quad (3.24)$$

where the degree of summability of $\tilde{u}_{i,t}$ is smaller than $S(1)$. In fact, this suggests an alternative two-step Pooled NLS estimator for Model (3.20) but not as efficient as the ordinary Pooled NLS estimator.

Corollary 3.6. *Let Assumptions 3.1, 3.2 and 3.4 hold respectively for $x_{i,t}$, $z_{i,t}$, $u_{i,t}$ and f_k with $k = 1, 2, 3, 4$. Moreover, if*

$$\begin{aligned} & \int_{-\infty}^{\infty} \beta_0^2 [1\{s \leq 0\} \frac{1}{1 + \exp(-\gamma_0(s - c_0))} - 1\{s \geq 0\} \frac{1}{1 + \exp(\gamma_0(s - c_0))} \\ & - \left(1\{s \leq 0\} \frac{1}{1 + \exp(-\gamma(s - c))} - 1\{s \geq 0\} \frac{1}{1 + \exp(\gamma(s - c))} \right)^2 ds > 0, \end{aligned}$$

for all $(\gamma, c) \neq (\gamma_0, c_0)$, we then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$\begin{aligned} R^*(\theta, \theta_0) &= \int_{-\infty}^{\infty} \beta_0^2 [1\{s \leq 0\} \frac{1}{1 + \exp(-\gamma_0(s - c_0))} - 1\{s \geq 0\} \frac{1}{1 + \exp(\gamma_0(s - c_0))} \\ &- \left(1\{s \leq 0\} \frac{1}{1 + \exp(-\gamma(s - c))} - 1\{s \geq 0\} \frac{1}{1 + \exp(\gamma(s - c))} \right)^2 ds \mathbf{E} \left[\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right], \end{aligned}$$

with $M_{n,T} = nT^{3/2}$. Moreover, for the Pooled NLS estimators $\hat{\gamma}_{n,T}$ and $\hat{c}_{n,T}$:

$$\hat{\gamma}_{n,T} \xrightarrow{p} \gamma_0,$$

$$\hat{c}_{n,T} \xrightarrow{p} c_0.$$

The strict inequalities appearing in Corollaries 3.5 and 3.6 are required by Condition (b) of Lemma 3.3 to identify the parameters. Furthermore, they are no more than special cases of what are required by Theorems 3.7 and 3.8. Next, the following corollary shows the asymptotic distribution of the Pooled NLS estimator,

Corollary 3.7. *Let Assumptions 3.1', 3.2 and 3.4 hold respectively for $x_{i,t}$, $z_{i,t}$, $u_{i,t}$ and f_k with $k = 1, 2, 3, 4$. Moreover, if the assumptions of Corollaries 3.5 and 3.6 hold, we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$(a) \sqrt{n}T \left[\begin{pmatrix} \hat{\alpha}_{n,T} \\ \hat{\beta}_{n,T} \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right] \xrightarrow{d} N(0, \Sigma_{\alpha_0, \beta_0}^{-1}),$$

$$(b) \sqrt{n} \sqrt[3/4]{T} \left[\begin{pmatrix} \hat{\gamma}_{n,T} \\ \hat{c}_{n,T} \end{pmatrix} - \begin{pmatrix} \gamma_0 \\ c_0 \end{pmatrix} \right] \xrightarrow{d} N(0, \Sigma_{\gamma_0, c_0}^{-1}),$$

where

$$\Sigma_{\alpha_0, \beta_0} = \mathbf{E} \left[\int_0^1 \begin{bmatrix} B_{x,i} \\ B_{x,i} 1\{B_{z,i} > 0\} \end{bmatrix} \begin{bmatrix} B_{x,i} \\ B_{x,i} 1\{B_{z,i} > 0\} \end{bmatrix}' dr \right],$$

and

$$\Sigma_{\gamma_0, c_0} = \int_{-\infty}^{\infty} \beta_0^2 \begin{bmatrix} \dot{f}_{\gamma_0}(s; \gamma_0, c_0) \\ \dot{f}_{c_0}(s; \gamma_0, c_0) \end{bmatrix} \begin{bmatrix} \dot{f}_{\gamma_0}(s; \gamma_0, c_0) \\ \dot{f}_{c_0}(s; \gamma_0, c_0) \end{bmatrix}' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right],$$

with \dot{f}_{γ_0} and \dot{f}_{c_0} being the first order partial derivatives of the transition function with respect to γ and c .

It can be verified that both \dot{f}_{γ_0} and \dot{f}_{c_0} satisfy Assumption I. Similar to the linear case, the above result (a) only holds for strict exogenous $x_{i,t}$ and $z_{i,t}$. Otherwise, $\hat{\alpha}_{n,T}$ and $\hat{\beta}_{n,T}$ are still consistent but the modification is necessary to guarantee $\sqrt{n}T$ rate of consistency. However, $\sqrt{n} \sqrt[3/4]{T}$ is the rate of consistency of $\hat{\gamma}_{n,T}$ and $\hat{c}_{n,T}$ in the presence of weak exogeneity.

3.5.2 Estimation of the Covariance Matrix

In this section, we consider the following five consistent estimators for the covariance matrix of $\hat{\theta}_{n,T}$ with $\hat{\theta}_{n,T} = (\hat{\alpha}_{n,T}, \hat{\beta}_{n,T}, \hat{\gamma}_{n,T}, \hat{c}_{n,T})'$ in which the second one is based on the two-step Pooled NLS estimator. We define

$$\hat{\Sigma}_{1,n,T} = \hat{\Sigma}_{n,T},$$

where $\hat{\Sigma}_{n,T}$ is defined in (3.13), and

$$\hat{\Sigma}_{2,n,T} = C_{n,T}^{-1} \begin{bmatrix} \sum_i \sum_t x_{i,t}^2 & \sum_i \sum_t x_{i,t}^2 1\{z_{i,t} \geq 0\} & 0 & 0 \\ \sum_i \sum_t x_{i,t}^2 1\{z_{i,t} \geq 0\} & \sum_i \sum_t x_{i,t}^2 1\{z_{i,t} \geq 0\} & 0 & 0 \\ 0 & 0 & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} \\ 0 & 0 & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{c}_{n,t}^2} \end{bmatrix} C_{n,T}^{-1},$$

where $\hat{Q}_{n,T}$ is the objective function which is evaluated at the estimates, and

$$C_{n,T} = \begin{bmatrix} T\sqrt{n} & 0 & 0 & 0 \\ 0 & T\sqrt{n} & 0 & 0 \\ 0 & 0 & T^{3/4}\sqrt{n} & 0 \\ 0 & 0 & 0 & T^{3/4}\sqrt{n} \end{bmatrix}.$$

Furthermore, we consider three other estimators that have the same asymptotic limit as those of $\hat{\Sigma}_{1,n,T}$ and $\hat{\Sigma}_{2,n,T}$. We define $\hat{\Sigma}_{3,n,T}$ as $\hat{\Sigma}_{1,n,T}$ excluding the correlation terms between $(\alpha, \beta)'$ and $(\gamma, c)'$, $\hat{\Sigma}_{4,n,T}$ as $\hat{\Sigma}_{1,n,T}$ excluding the second order derivatives of the transition function with respect to γ and c , and $\hat{\Sigma}_{5,n,T}$ as $\hat{\Sigma}_{1,n,T}$ excluding both the second derivatives and the cross correlation terms.

Corollary 3.8. *Let Assumptions 3.1', 3.2 and 3.4 hold respectively for $x_{i,t}$, $z_{i,t}$, $u_{i,t}$ and f_k with $k = 1, 2, 3, 4$. Moreover, if the assumptions of Corollaries 3.5 and 3.6 hold, we then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{l,n,T} \xrightarrow{p} \Sigma_{STR},$$

for $l = 1, 2, 3, 4, 5$, where

$$\Sigma_{STR} = \begin{bmatrix} \Sigma_{\alpha_0, \beta_0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\gamma_0, c_0} \end{bmatrix},$$

with \dot{f}_{γ_0} and \dot{f}_{c_0} being the first order partial derivative of the transition function with respect to γ and c .

3.6 Monte Carlo Simulations

In this section, we consider the finite sample accuracy of the theoretical results in the case of panel smooth transition co-summable regressions. We perform 5000 simulations with different combinations of $n = 5, 10, 30, 60, 90, 120$ and $T = 30, 60, 90, 120, 150$ to show the finite sample performance of the Pooled NLS estimator. The chosen combinations of the sample size fit most applications in macroeconomic studies where the cross-sectional dimension is not too large. If the data is monthly collected, a sample of size $T = 150$ contains more than 10 years data. We take Model (3.20) and set the values of the parameters as $\alpha_0 = -1$, $\beta_0 = 1.5$, $\gamma_0 = 1$ and $c_0 = 0.5$. The starting values for the Newton algorithm are chosen as $(-0.5, 0.8, 0.5, 0)$. We only consider the homogeneous panel case

with strongly exogenous regressors². For the data generating process, we set, for all i ,

$$\Sigma_i = \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and

$$\varphi_{i,t} = \rho_{i,t} \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}, \quad (3.25)$$

where $\rho_{i,t}$ is chosen from a uniform distribution $[0.77, 1.98]$. The innovation terms are generated by an AR(1) process with coefficient $\varphi_{i,t}$. By doing so, we make sure that a unit root is excluded and the time variation is not reduced by $\rho_{i,t}$. $y_{i,t}$ is generated using the smooth transition co-summable regression Model (3.20). In summary, the DGP is as the following:

$$y_{i,t} = (-1) \times x_{i,t} + 1.5 \times \frac{1}{1 + \exp(-1 \times (z_{i,t} - 0.5))} x_{i,t} + u_{i,t}, \quad (3.26)$$

where $u_{i,t}$ is generated by IID standard *Normal* random variables for each unit i . Moreover, $z_{i,t}$ and $x_{i,t}$ are scalars of regressors assumed to be generated by

$$\begin{bmatrix} x_{i,t} \\ z_{i,t} \end{bmatrix} = \begin{bmatrix} x_{i,t-1} \\ z_{i,t-1} \end{bmatrix} + \begin{bmatrix} v_{x,i,t} \\ v_{z,i,t} \end{bmatrix}, \quad (3.27)$$

with common initialization at $t = 0$ satisfying

$$(z_{i,0}, x_{i,0})' = (0, 0)'.$$

The innovation term $V_{i,t} = (v_{x,i,t}, v_{z,i,t})'$ is generated by

$$V_{i,t} = \varphi_{i,t} V_{i,t-1} + \eta_{i,t}, \quad (3.28)$$

where $\eta_{i,t}$ is generated by IID standard *Normal* random variables for each unit i .

²The consistency result also holds for the case with weakly exogenous regressors for which results are available upon request.

Tables 3.1-3.4 show the mean absolute error(MAE) that is defined as

$$MAE = \frac{\sum_{j=1}^{ns} |\hat{\theta}_{n,T} - \theta|}{ns},$$

where $\hat{\theta}_{n,T}$ is the Pooled NLS estimates, for different n and T dimensions. The results coincide with the asymptotic properties that are derived in Section 3.5. First, for any given n , the MAE decreases as T increases. Besides, by moving from the left-upper corner to the right-lower corner of the table, the MAE decreases as n and T jointly increase, which indicates the possible joint consistency of the Pooled NLS estimator. Second, comparing the results for different parameters, the MAEs of $\hat{\alpha}_{n,T}$ and $\hat{\beta}_{n,T}$ are smaller than those of $\hat{\gamma}_{n,T}$ and $\hat{c}_{n,T}$ with the same sample size. This is an indication that the speeds of convergence for the former two are faster than those for the later two.

Figures 3.1-3.8 show the estimated density function of the following transformation of the estimator

$$\sqrt{\hat{\Sigma}_{l,n,T}} C_{n,T}(\hat{\theta}_{n,T} - \theta_0), \quad (3.29)$$

with $l = 1, 2$. We only present the empirical probability density functions of the Pooled NLS estimator for cross-sectionally independent panel co-summability with strictly exogenous regressors and $n = 10$ because the performance is good given such a small cross-sectional dimension³. Namely, a kernel estimator is applied to the density functions of the 5000 replications. Figures 1, 3, 5 and 7 show that, based on the estimator $\hat{\Sigma}_{1,n,T}$, the estimated probability density function is close to that of the standard normal distribution, with sample size $n = 10$ and $T = 30, 60, 90, 120, 150$. Considering the possible distortion caused by the bias of the kernel estimation, the finite sample performance of the Pooled NLS estimator is close to its asymptotic distribution. It is worth of noticing that, for $n = 5$, the kernel estimation of the density function does not perform well in the presence of outliers in the estimates of γ_0 and c_0 . Regarding the outliers, we cannot distinguish the effect due to the algorithm used for the nonlinear optimization from that due to the probabilistic properties of the Pooled NLS estimator. Besides, we include the results of $\hat{\Sigma}_{2,n,T}$ for comparison. First, the results show that no matter which estimator is chosen for the covariance matrix, the finite sample distribution of (3.29) is close to the standard normal distribution. Second, ignoring the second derivative terms does not cause much distortion from the standard normal distribution. Namely, $\hat{\Sigma}_{1,n,T}$ and $\hat{\Sigma}_{4,n,T}$ are preferred in practice when T is small. Furthermore, the remaining tables show the sample covariances for the

³The results for $n = 30, 60, 90, 120$ are similar and available upon request.

5000 replications of the transformed estimates by (3.29). They are very close to the ideal value 1, the standard deviation of a standard *Normal* distribution, which is another confirmation of the good finite sample performance of the Pooled NLS estimator.

Intuitively, the logistic STR co-summable regression model used in the above simulation could be interpreted as regimes switching processes with smooth and continuous transition between regimes. Hence, it is of no surprise if the realizations of the logistic STR co-summable model behavior like piecewise linear models. In fact, the following simulation results show that the realization of the logistic STR co-summable model can behave very differently from a piecewise linear model. Moreover, the performance of the Pooled NLS estimator is independent from the behaviors of the realizations of the logistic STR co-summable model. Figure 3.9 shows four realizations generated based on the setup specified till Eq. 3.28. For each of the four realization, $y_{i,t}$ on the vertical axis is plotted against $x_{i,t}$ on the horizontal axis with $T = 30$ by $n = 10$ and totally 300 observations being pooled together. The piecewise linearity is very obvious as been shown in the plot. In view of (3.24), (3.28) can be approximated by the following:

$$y_{i,t} = (-1) \times x_{i,t} + 1.5 \times \frac{1}{1 + \exp(-1 \times (z_{i,t} - 0.5))} x_{i,t} + u_{i,t}.$$

Moreover, according to (3.25), $x_{i,t}$ and $z_{i,t}$ are positively correlated. Hence, (3.24) implies that if $z_{i,t} > 0$, then $x_{i,t} > 0$ being right to the origin with

$$y_{i,t} \approx (\alpha + \beta)x_{i,t} = 0.5x_{i,t}.$$

On the other hand, if $z_{i,t} < 0$, then $x_{i,t} < 0$ being left to the origin with

$$y_{i,t} \approx \alpha x_{i,t} = -1x_{i,t}.$$

The noise around the origin is due to the observations generated by the transition stage. Consequently, if we set

$$\varphi_{i,t} = \rho_{i,t} \begin{bmatrix} -0.3 & -0.2 \\ 0.2 & 0.3 \end{bmatrix}, \quad (3.30)$$

the pattern opposite to what have been observed in Figure 3.9 is expected. Indeed, Figure 3.10 confirms the expectation when $x_{i,t}$ and $z_{i,t}$ are negatively correlated. Therefore, if $z_{i,t} > 0$, then

$x_{i,t} < 0$ with

$$y_{i,t} \approx (\alpha + \beta)x_{i,t} = 0.5x_{i,t}.$$

On the other hand, if $z_{i,t} < 0$, then $x_{i,t} > 0$ with

$$y_{i,t} \approx \alpha x_{i,t} = -1x_{i,t}.$$

The noise due to the transition stage is again around the origin. In addition, if $x_{i,t}$ and $z_{i,t}$ are independent, the above discussion does not hold in general and, according to (3.24), no piecewise linear pattern is expected in the realization of Eq. 3.28. Figure 3.11 illustrates four realizations of this case with

$$\varphi_{i,t} = \rho_{i,t} \begin{bmatrix} -0.4 & 0 \\ 0 & -0.4 \end{bmatrix}. \quad (3.31)$$

For completeness, the following tables and figures show the robustness of the Pooled NLS estimator against realizations without piecewise linear properties. Tables 3.10-3.13 show the MAE for the Pooled NLS estimates, for different n and T dimensions. The setup of the simulation is exactly the same as that at the beginning of this section only with (3.25) being replaced by (3.31). The results are in line with the asymptotic properties that are derived in Section 3.5. Figures 3.12-3.15 show the estimated density function of the transformation of the estimator, which is defined by (3.29), with $l = 1$. The overall conclusion is that the performance of the Pooled NLS estimator is robust against the piecewise linear or nonlinear patterns observed in the realizations of Model 3.26 in finite samples.

3.7 Conclusion

In this chapter, we apply the asymptotic theory of nonlinear nonstationary time series to obtain the asymptotic results of estimators for nonlinear nonstationary panel data models. In the presence of both nonlinearity and nonstationarity, the definitions of summability and co-summability introduced by Rico (2009) are adopted to interpret the balance of the model equation. Different nonlinear transformations of nonstationary time series may result in different orders of summability. Hence, we distinguish the simple strong co-summability from multiple strong co-summability situations according to the different nonlinear functions that are parts of the model specification. The latter case encompasses the linear cointegration models with short run dynamics. The general setup

considered in Sections 3.3 and 3.4 encompasses threshold cointegration models, polynomial functions of $I(1)$ processes and time-varying coefficients models. A Pooled Nonlinear Least Squares estimator is proposed for a homogeneous panel with independent cross-section units with respect to simple strong co-summability and multiple strong co-summability respectively. The estimator is proved to be consistent and asymptotically normally distributed when $T \rightarrow \infty$ followed by $n \rightarrow \infty$. If the nonlinear transformation is I -regular, the rate of convergence is $T^{3/4}\sqrt{n}$ and weak exogeneity is sufficient for the asymptotic distribution. While for the H -regular functions, the rate is $T\sqrt{n}$ and strict exogeneity is required for the asymptotic normality and freedom from nuisance parameters. Furthermore, conditions for the existence of the estimator and for identification are introduced. The consistency results are derived based on the asymptotic properties of the objective functions. In the case of multiple strong co-summability, the derivation of consistency implies a consistent k -stage Pooled NLS estimator which is not necessarily equivalent to the one-shot Pooled NLS estimator, and may have a different rate of consistency and different asymptotic distributions.

The simulation study is set for the Logistic Smooth Transition Regression specification. The reparametrization shows that it is actually a special case of the multiple strong co-summable regression models. The simulation results show that the Pooled NLS estimator performs well in finite samples. We do not find any difficulty with identifying the parameters. One possible explanation would be that under the panel set up, the limit of the objective function is deterministic. Hence, the identification conditions are almost surely satisfied. While in the time series case, the limit of the objective function is a functional of Brownian motion, which might imply a positive probability of failing the identification conditions. Finally, several estimators are proposed for the covariance matrix. The simulation results show that some of them perform better than the others when T is small. In any case, when T increases, the finite sample distribution of the constructed statistic is very close to the standard Normal distribution.

It is finally worth noting some limitations and possible extensions. First, the cross-sectional independence is critical for all above conclusions to be valid. However, this assumption is not regarded to be in line with reality. There might be several ways to relax this restriction such as introducing a factor structure. Second, in practice, regressors are very often endogenous. Hence, in the presence of endogeneity, a modification is required for the inference purpose just as the linear case. We notice that the panel model specified by polynomial functions of $I(1)$ time series with endogenous regressors are developed by Hong and Wagner (2008) and Hong and Wagner (2011). Moreover, the time series results in Saikkonen and Choi (2004), De Jong (2005) and Kasparis (2008) can be extended to the

nonlinear panel models with H – *regular* transformations of endogenous $I(1)$ regressors. Third, our simulation study indicates that under cross section independence the Pooled NLS estimator is also consistent as T and n jointly diverge to infinity. However, its theoretical derivation requires further research. Besides, the K –stage Pooled NLS estimator is only proved to be consistent at this stage. Just as the issue of endogeneity, its asymptotic distribution requires further study. Further extensions of the model includes adding deterministic components for individual cross-sectional units, nonlinear transformations of stationary components, heterogeneity and so on.

3.A Tables

Table 3.1: MAE for Pooled NLS: $\hat{\alpha}_{n,T}$

N \ T	30	60	90	120	150
5	0.0241	0.0088	0.0056	0.0041	0.0031
10	0.0102	0.0045	0.0029	0.0021	0.0017
30	0.0049	0.0023	0.0015	0.0011	0.0009
60	0.0034	0.0016	0.0010	0.0007	0.0006
90	0.0027	0.0013	0.0008	0.0006	0.0005
120	0.0023	0.0011	0.0007	0.0005	0.0004

Table 3.2: MAE for Pooled NLS: $\hat{\beta}_{n,T}$

N \ T	30	60	90	120	150
5	0.0409	0.0149	0.0090	0.0064	0.0048
10	0.0160	0.0068	0.0045	0.0032	0.0025
30	0.0074	0.0034	0.0021	0.0016	0.0012
60	0.0050	0.0023	0.0015	0.0011	0.0009
90	0.0040	0.0018	0.0012	0.0009	0.0007
120	0.0034	0.0016	0.0010	0.0007	0.0006

Table 3.3: MAE for Pooled NLS: $\hat{\gamma}_{n,T}$

N \ T	30	60	90	120	150
5	0.1583	0.0841	0.0544	0.0404	0.0347
10	0.0704	0.0408	0.0291	0.0238	0.0201
30	0.0351	0.0203	0.0150	0.0122	0.0102
60	0.0236	0.0138	0.0104	0.0083	0.0071
90	0.0191	0.0114	0.0082	0.0067	0.0057
120	0.0165	0.0098	0.0070	0.0058	0.0049

Table 3.4: MAE for Pooled NLS: $\hat{c}_{n,T}$

N \ T	30	60	90	120	150
5	0.1515	0.0884	0.0615	0.0488	0.0411
10	0.0816	0.0472	0.0343	0.0278	0.0233
30	0.0397	0.0231	0.0172	0.0140	0.0118
60	0.0273	0.0161	0.0117	0.0094	0.0082
90	0.0217	0.0130	0.0096	0.0078	0.0065
120	0.0191	0.0113	0.0082	0.0066	0.0056

Table 3.5: Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{1,n,T}$

T \ <i>Trans</i>	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	0.9869	1.0063	1.0581	0.9997
60	1.0029	0.9793	1.0359	1.0154
90	0.9925	1.0208	1.0042	0.9930
120	0.9496	1.0074	1.0053	1.0060
150	1.0051	1.0145	1.0023	0.9756

Table 3.6: Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{2,n,T}$

T \ <i>Trans</i>	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.2028	1.5075	1.3206	1.2529
60	1.1069	1.2255	1.1650	1.1511
90	1.0553	1.2225	1.0963	1.0913
120	1.0121	1.1457	1.0797	1.0838
150	1.0559	1.1232	1.0639	1.0437

Table 3.7: Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{3,n,T}$

T \ <i>Trans</i>	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.1948	1.3069	1.3169	1.2514
60	1.1043	1.1273	1.1654	1.1508
90	1.0552	1.1467	1.0945	1.0913
120	1.0114	1.0869	1.0788	1.0838
150	1.0549	1.0745	1.0647	1.0435

Table 3.8: Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{4,n,T}$

T \ <i>Trans</i>	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	0.9869	1.0063	1.0614	1.0016
60	1.0029	0.9793	1.0355	1.0157
90	0.9925	1.0208	1.0059	0.9931
120	0.9496	1.0074	1.0063	1.006
150	1.0051	1.0145	1.0015	0.9758

Table 3.9: Estimated Variance for Pooled NLS: $n = 10$, $\hat{\Sigma}_{5,n,T}$

T \ <i>Trans</i>	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.1948	1.3069	1.3206	1.2529
60	1.1043	1.1273	1.1650	1.1511
90	1.0552	1.1467	1.0963	1.0913
120	1.0114	1.0869	1.0797	1.0838
150	1.0549	1.0745	1.0639	1.0437

Table 3.10: MAE for Pooled NLS: $\hat{\alpha}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.1531	0.0426	0.0256	0.0183	0.0134
10	0.0571	0.0229	0.0141	0.0100	0.0077
30	0.0281	0.0114	0.0071	0.0052	0.0039
60	0.0184	0.0079	0.0048	0.0035	0.0027
90	0.0151	0.0065	0.0040	0.0028	0.0022
120	0.0129	0.0054	0.0034	0.0025	0.0019

Table 3.11: MAE for Pooled NLS: $\hat{\beta}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.4096	0.0930	0.0503	0.0348	0.0247
10	0.1202	0.0415	0.0244	0.0172	0.0129
30	0.0519	0.0198	0.0116	0.0085	0.0063
60	0.0328	0.0133	0.0081	0.0055	0.0042
90	0.0270	0.0108	0.0064	0.0045	0.0034
120	0.0230	0.0091	0.0055	0.0039	0.0030

Table 3.12: MAE for Pooled NLS: $\hat{\gamma}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.3728	0.1332	0.0833	0.0619	0.0502
10	0.1586	0.0716	0.0480	0.0368	0.0300
30	0.0796	0.0375	0.0245	0.0193	0.0158
60	0.0524	0.0251	0.0170	0.0131	0.0106
90	0.0421	0.0207	0.0139	0.0108	0.0086
120	0.0360	0.0175	0.0119	0.0093	0.0075

Table 3.13: MAE for Pooled NLS: $\hat{c}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.4695	0.1594	0.1045	0.0768	0.0614
10	0.1832	0.0807	0.0568	0.0433	0.0349
30	0.0846	0.0418	0.0281	0.0218	0.0179
60	0.0573	0.0283	0.0192	0.0150	0.0123
90	0.0458	0.0230	0.0152	0.0121	0.0099
120	0.0393	0.0199	0.0134	0.0106	0.0084

3.B Figures

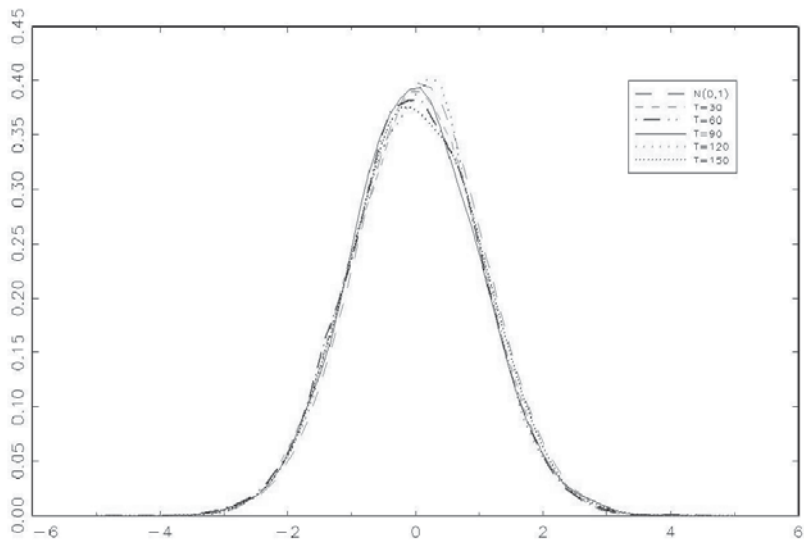


Figure 3.1: Estimated PDF for Pooled NLS: $n = 10, \alpha, \hat{\Sigma}_{1,n,T}$

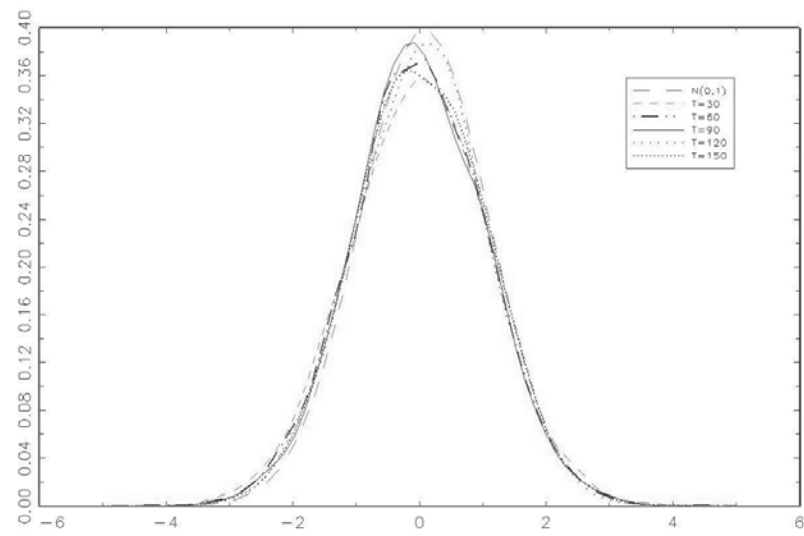


Figure 3.2: Estimated PDF for Pooled NLS: $n = 10, \alpha, \hat{\Sigma}_{2,n,T}$

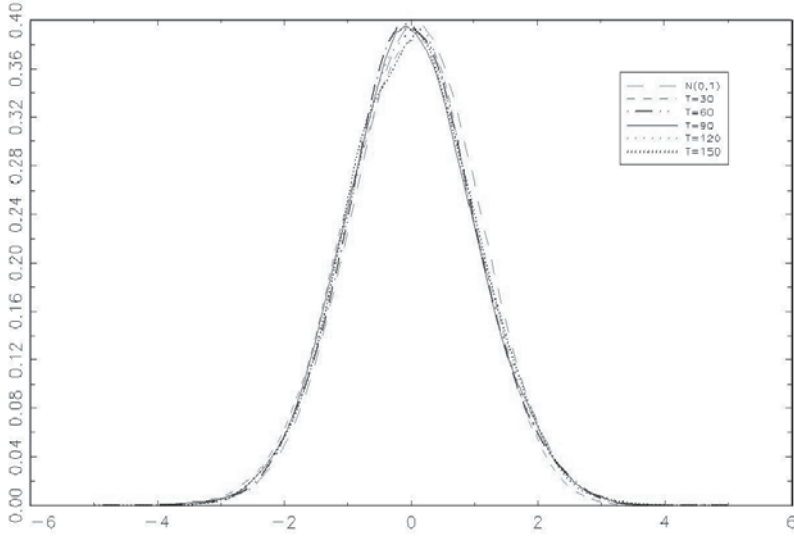


Figure 3.3: Estimated PDF for Pooled NLS: $n = 10$, $\beta, \hat{\Sigma}_{1,n,T}$

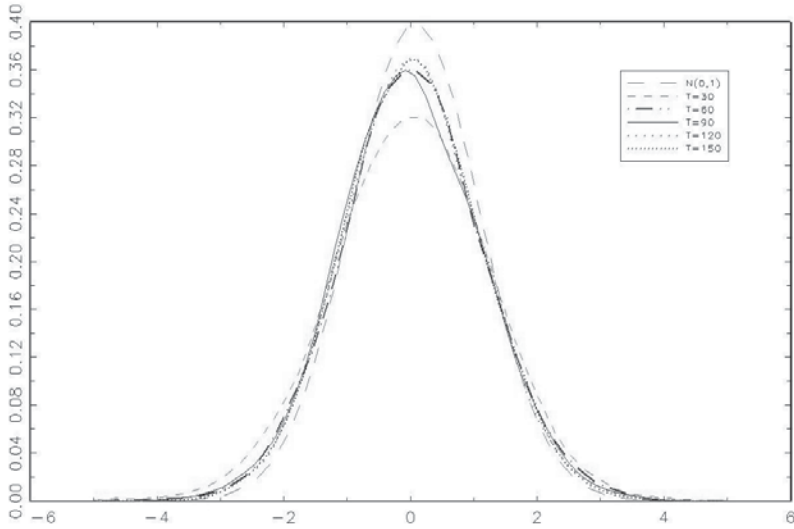


Figure 3.4: Estimated PDF for Pooled NLS: $n = 10$, $\beta, \hat{\Sigma}_{2,n,T}$

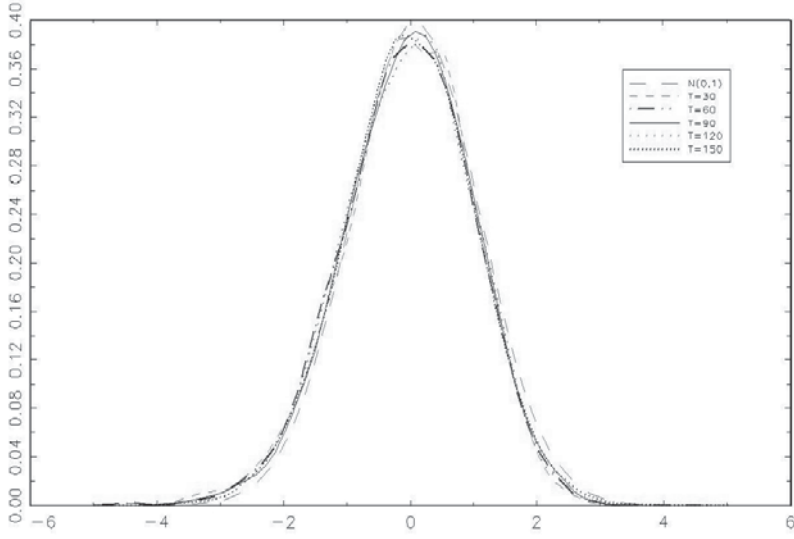


Figure 3.5: Estimated PDF for Pooled NLS: $n = 10$, γ , $\hat{\Sigma}_{1,n,T}$

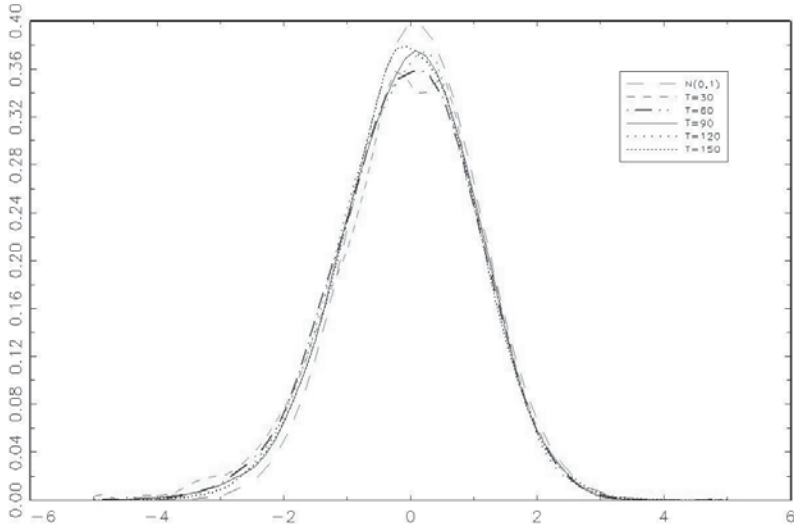


Figure 3.6: Estimated PDF for Pooled NLS: $n = 10$, γ , $\hat{\Sigma}_{2,n,T}$

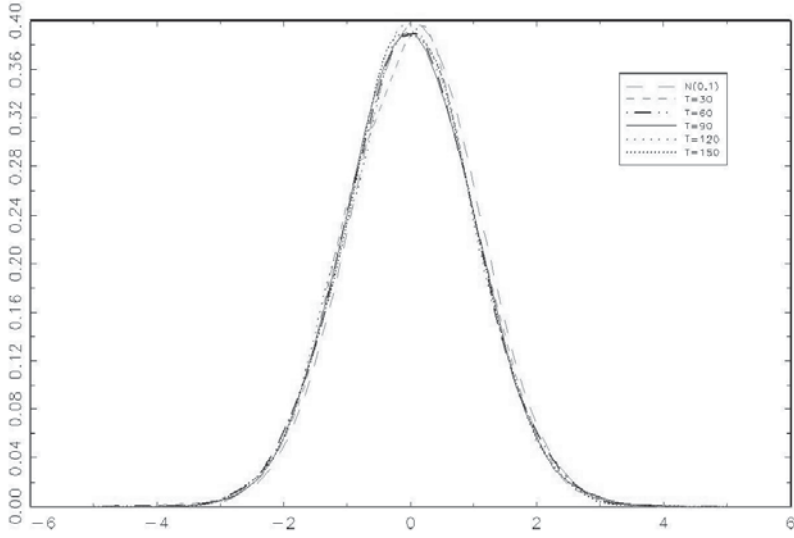


Figure 3.7: Estimated PDF for Pooled NLS: $n = 10$, c , $\hat{\Sigma}_{1,n,T}$

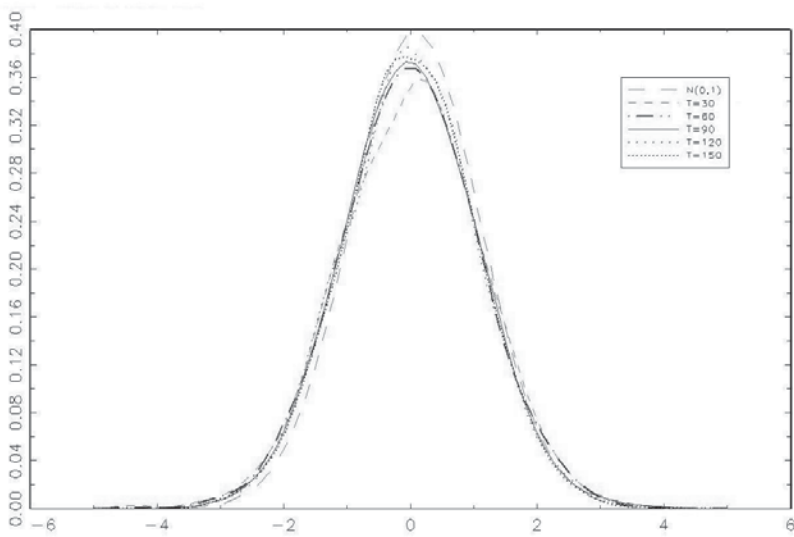


Figure 3.8: Estimated PDF for Pooled NLS: $n = 10$, c , $\hat{\Sigma}_{2,n,T}$

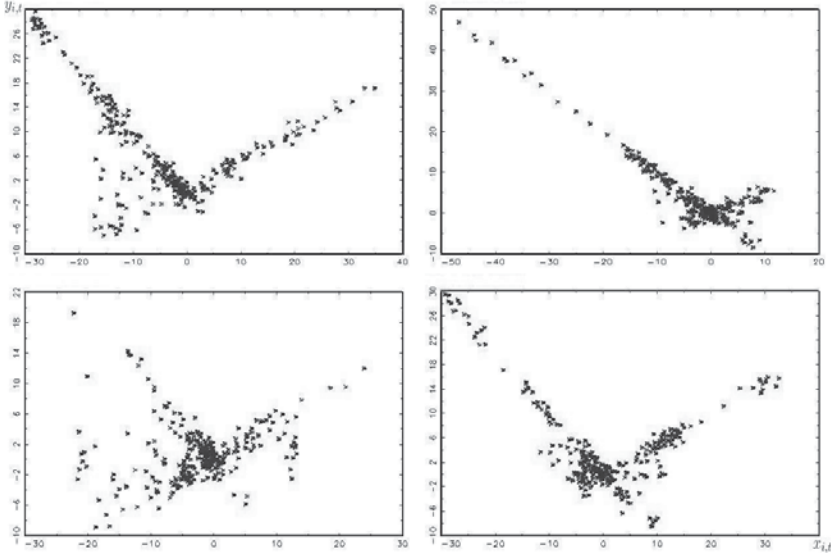


Figure 3.9: Realizations: $n = 10$, $T = 30$, $COV(x_{i,t}, z_{i,t}) > 0$

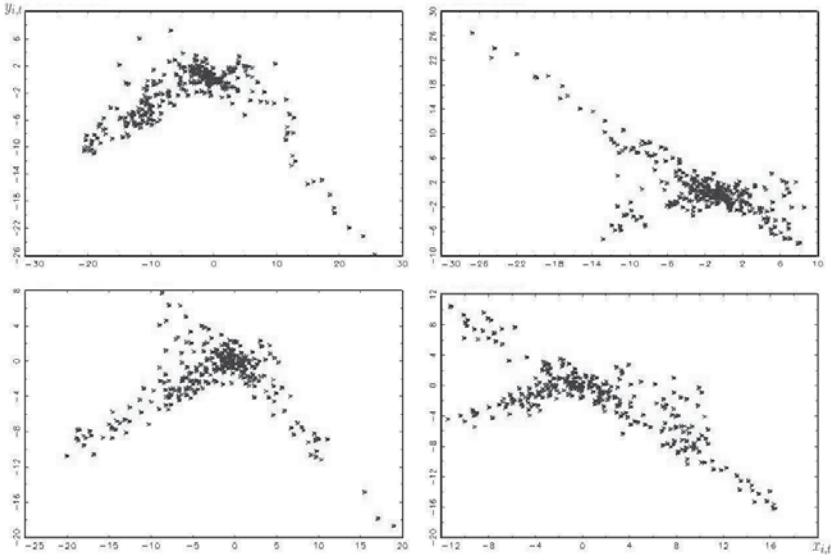


Figure 3.10: Realizations: $n = 10$, $T = 30$, $COV(x_{i,t}, z_{i,t}) < 0$

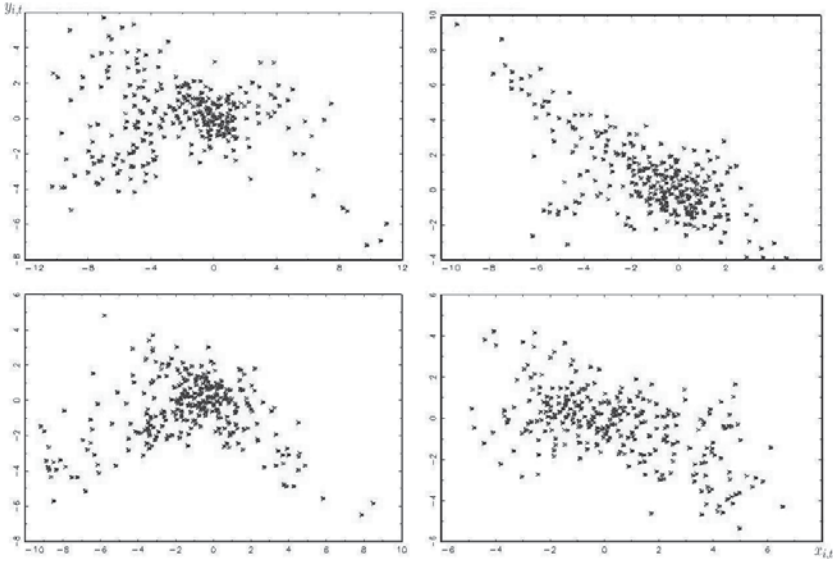


Figure 3.11: Realizations: $n = 10$, $T = 30$, $\text{COV}(x_{i,t}, z_{i,t}) = 0$

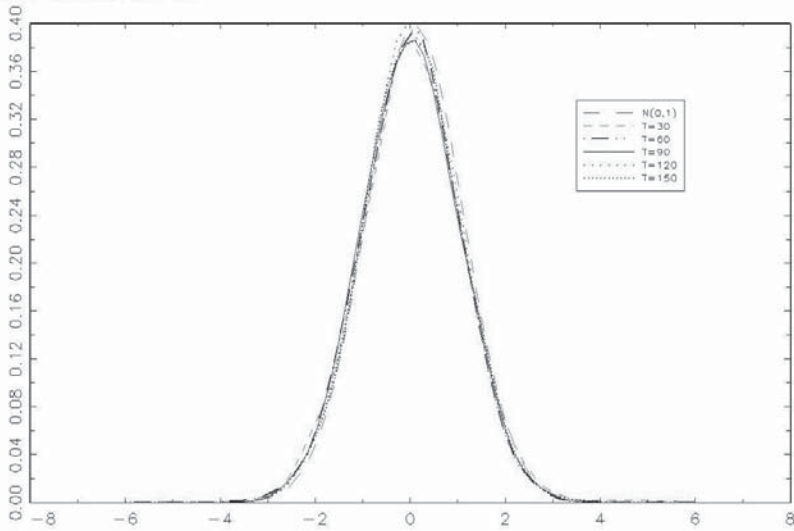


Figure 3.12: Estimated PDF for Pooled NLS: $n = 10$, α , $\hat{\Sigma}_{1,n,T}$, $\text{COV}(x_{i,t}, z_{i,t}) = 0$

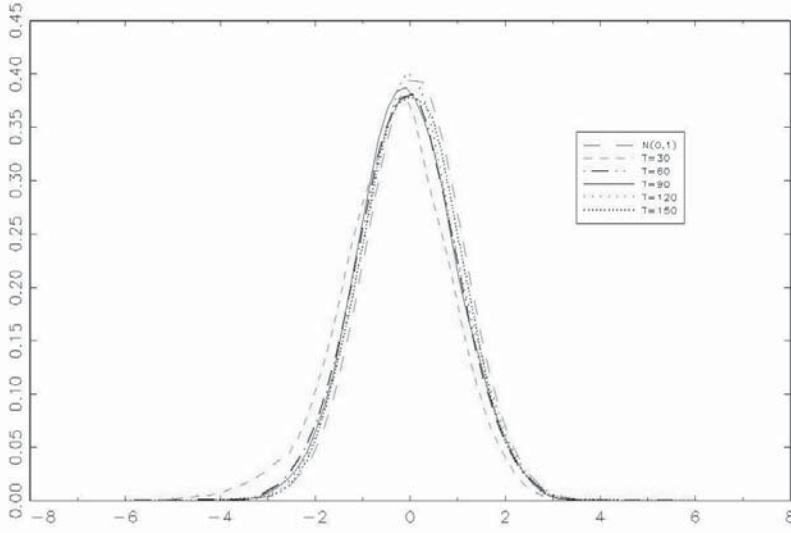


Figure 3.13: Estimated PDF for Pooled NLS: $n = 10$, β , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

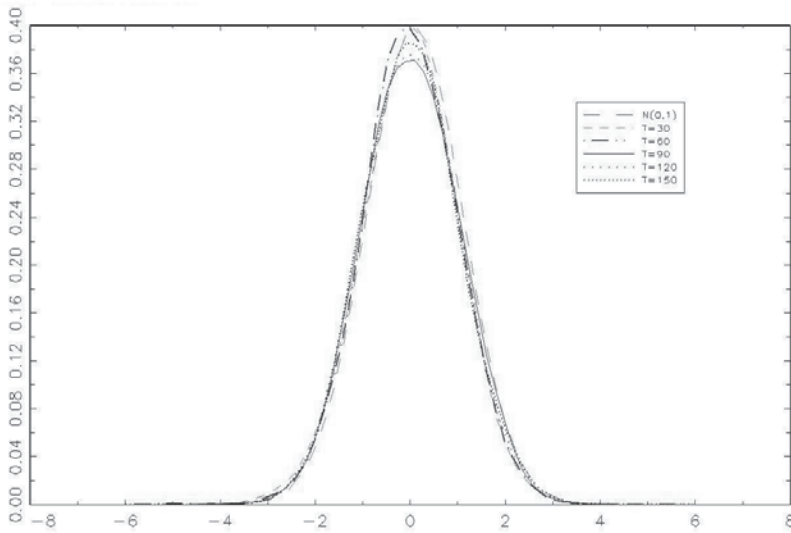


Figure 3.14: Estimated PDF for Pooled NLS: $n = 10$, γ , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

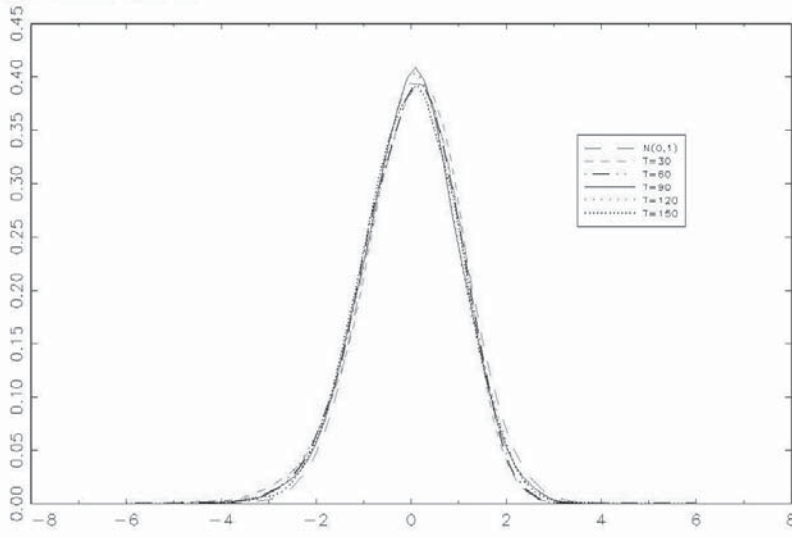


Figure 3.15: Estimated PDF for Pooled NLS: $n = 10$, c , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

3.C Appendix

The proof starts with some simplification. First, the sequence $\{(z_{i,t}, x_{i,t})'\}_{t=0}^\infty$ may be initialized at $t = 0$ with $(z_{i,0}, x_{i,0})' = O_p(1)$. However, in the proofs of the lemmas and the theorems, it is set $z_{i,0} = x_{i,0} = 0$ to avoid any unnecessary complication in the exposition. Just by adding a term $x_{i,0}/\sqrt{nT^\epsilon}$ or $z_{i,0}/\sqrt{nT^\epsilon}$ with $\epsilon > 0$, the results of the FCLT will not change since $x_{i,0}/\sqrt{nT^\epsilon} \rightarrow 0$ and $z_{i,0}/\sqrt{nT^\epsilon} \rightarrow 0$. Second, assume $\mathbf{E}u_{i,t}^2 = 1$. Other values simply have a scaling effect in the subsequent analysis of this section.

3.C.1 Lemma 3.C.1

Lemma 3.C.1. : Suppose that $x_{i,t}$, $z_{i,t}$ and $u_{i,t}$ satisfy Assumptions 3.1-3.2 with $z_{i,t}$ and $x_{i,t}$ being integrated time series defined in (3.2) and (3.3). Let $f: \mathbb{R} \rightarrow \mathbb{R}$ is I -regular. Then, the following holds, as $T \rightarrow \infty$:

$$\frac{1}{(\sqrt{T})^{(1+k)}} \sum_{t=1}^T x_{i,t}^k f(z_{i,t}) \xrightarrow{p} \int_{-\infty}^{\infty} f(s) ds \int_0^1 B_{x,i}^k(r) dL_{z,i}(r, 0),$$

and

$$\frac{1}{(\sqrt{T})^{(1/2+k)}} \sum_{t=1}^T x_{i,t}^k f(z_{i,t}) u_{i,t} \xrightarrow{d} \left(\int_{-\infty}^{\infty} f(s)^2 ds \int_0^1 B_{x,i}^{2k}(r) dL_{z,i}(r, 0) \right)^{1/2} W_i(1),$$

with $k > 0$.

$B_{x,i}$, $B_{z,i}$, $L_{z,i}(r, 0)$ and W_i are the same as those defined in Lemma 3.1. as those in a common probability space supporting $B_{u,i}$ and $B_{v,i}$. The proof is based on Theorem 5.1 in Park and Phillips (1999), Lemma 2.1 and Theorem 3.2 in Park and Phillips (2001). The main idea is to approximate the integrable function by a step function to determine the effect of local time at point 0. Then, we construct the convergence of the quadratic processes of a martingale difference process. Finally, we use martingale difference processes, its Dubins-Schwarz Brownian motion and stopping time to prove the result.

3.C.2 Proof of Lemma 3.C.1

Let $\kappa_f = T^a$ and $\delta_f = T^{-b}$ for $a, b > 0$ satisfying

$$a - 3b + 1 < 0, \tag{3.C.1}$$

$$(6b - 3)p - 3 + 4p > 0, \tag{3.C.2}$$

$$2a - b - 1 < 0, \tag{3.C.3}$$

$$(a - b)p - 1 > 0. \tag{3.C.4}$$

According to the Skorohod representation argument in Section 3.2, and assuming $x_{i,0} = z_{i,0} = 0$,

$$\frac{1}{(\sqrt{T})^{(1+k)}} \sum_{t=1}^T x_{i,t}^k f(z_{i,t}) = {}^d \sqrt{T} \int_0^1 f(\sqrt{T} B_{x,i,T}(r)) B_{z,i,T}^k(r) dr.$$

If $x_{i,0} = z_{i,0} \neq 0$, we simply replace $x_{i,t}^k f(z_{i,t})$ with $(x_{i,t} + x_{i,0})^k f(z_{i,t} + x_{i,0})$. All the proofs go through given $x_{i,0}/\sqrt{T} \xrightarrow{a.s.} 0$. Then, define f_T, f'_T and f''_T as

$$f_T(x) = f(x)1\{-\kappa_T\delta_T \leq x < \kappa_T\delta_T\},$$

$$f'_T(x) = f(x)1\{x \geq \kappa_T\delta_T\},$$

$$f''_T(x) = F(x)1\{x < -\kappa_T\delta_T\},$$

so that $f = f_T + f'_T + f''_T$. Now we need to prove

$$\sqrt{T} \int_0^1 f_T(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k(r)dr = \int_{-\infty}^{\infty} f(s)ds \int_0^1 B_{x,i}^k(r)dL_{z,i}(r,0) + o_p(1), \quad (3.C.5)$$

$$\sqrt{T} \int_0^1 f'_T(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k(r)dr = o_p(1), \quad (3.C.6)$$

$$\sqrt{T} \int_0^1 f''_T(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k(r)dr = o_p(1). \quad (3.C.7)$$

First, we prove (3.C.5):

Define a new function $f_{\delta_T}(x)$ as:

$$f_{\delta_T}(x) = \sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T)1\{k\delta_T \leq x < (k+1)\delta_T\}.$$

Then, by the Lipschitz condition given in Definition 3.3 of Park and Phillips (2001) with the Euclidean norm,

$$\begin{aligned} \sup_x |f_T(x) - f_{\delta_T}(x)| &\leq \sum_{k=-\kappa_T}^{\kappa_T-1} \sup_x |f_T(x) - f_T(k\delta_T)|1\{k\delta_T \leq x < (k+1)\delta_T\} \\ &\leq \sum_{k=-\kappa_T}^{\kappa_T-1} \sup_x c|x - k\delta_T|1\{k\delta_T \leq x < (k+1)\delta_T\} \\ &\leq \sum_{k=-\kappa_T}^{\kappa_T-1} c|(k+1)\delta_T - k\delta_T|1\{k\delta_T \leq x < (k+1)\delta_T\} \\ &\leq c\delta_T. \end{aligned}$$

The first inequality follows from the triangular inequality. Therefore,

$$\left| \sqrt{T} \int_0^1 f_T(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k(r)dr - \sqrt{T} \int_0^1 f_{\delta_T}(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k(r)dr \right|^2$$

$$\begin{aligned}
&= |\sqrt{T} \int_0^1 \sum_{k=-\kappa_T}^{\kappa_T-1} f(\sqrt{T}B_{z,i,T}(r)) 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\} B_{z,i,T}^k dr \\
&\quad - \sqrt{T} \int_0^1 \sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\} B_{x,i,T}^k dr|^2, \\
&\leq \left| \sqrt{T} \int_0^1 \sum_{k=-\kappa_T}^{\kappa_T-1} [f(\sqrt{T}B_{z,i,T}(r)) - f(k\delta_T)] 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\} B_{x,i,T}^k dr \right|^2,
\end{aligned}$$

by Hölder's inequality,

$$\leq \left\{ \sqrt{T} \int_0^1 \sum_{k=-\kappa_T}^{\kappa_T-1} |f(\sqrt{T}B_{z,i,T}(r)) - f(k\delta_T)| 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\} |B_{x,i,T}|^k dr \right\}^2,$$

by Lipschitz condition,

$$\leq (\sqrt{T})^2 \left\{ \int_0^1 \sum_{k=-\kappa_T}^{\kappa_T-1} c\delta_T 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\} |B_{x,i,T}|^k dr \right\}^2,$$

by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\leq (\sqrt{T})^2 c^2 \delta_T^2 \int_0^1 \left[\sum_{k=-\kappa_T}^{\kappa_T-1} 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\} \right]^2 dr \int_0^1 |B_{x,i,T}|^{2k} dr, \\
&= \sqrt{T} c^2 \delta_T^2 \left[\sqrt{T} \int_0^1 1\{-\kappa_T\delta_T \leq \sqrt{T}B_{z,i,T}(r) < \kappa_T\delta_T\} dr \right] \int_0^1 |B_{x,i,T}|^{2k} dr, \\
&= \sqrt{T} c^2 \delta_T^2 (\kappa_T\delta_T) \left[\frac{\sqrt{T}}{(\kappa_T\delta_T)} \int_0^1 1\left\{-\frac{\kappa_T\delta_T}{\sqrt{T}} \leq B_{z,i,T}(r) < \frac{\kappa_T\delta_T}{\sqrt{T}}\right\} dr \right] \int_0^1 |B_{x,i,T}|^{2k} dr, \\
&= (c^2 \sqrt{T} \delta_T^3 \kappa_T) \left[\frac{\sqrt{T}}{\kappa_T\delta_T} N_{z,i,T}(-\kappa_T\delta_T, \kappa_T\delta_T) \right] \int_0^1 |B_{x,i,T}|^{2k} dr, \\
&= O_p(\sqrt{T} \kappa_T \delta_T^3) = o_p(1).
\end{aligned}$$

Here, define $N_{z,i,T}(a, b) = N_{z,i,T}(\sqrt{T} : a, b)$ and $N_{z,i}(a, b) = N_{z,i}(\sqrt{T} : a, b)$ as that in Park and Phillips (1999). Following Lemma 2.3 of Park and Phillips (1999) and the continuous mapping theorem, $\int_0^1 |B_{x,i,T}|^{2k} dr \xrightarrow{p} \int_0^1 |B_{x,i}|^{2k} dr$ is well defined because $|B_{x,i}|$ is half-normal. From (3.C.2) and Lemma 2.5(b) of Park and Phillips (1999), it follows that,

$$\frac{\sqrt{T}}{\kappa_T\delta_T} N_{z,i,T}(-\kappa_T\delta_T, \kappa_T\delta_T) = 2L_{z,i}(1, 0) + o_p(1).$$

Then, the last equality is due to (3.C.1) and the definitions of κ_T and δ_T . Now, we have

$$\sqrt{T} \int_0^1 f_T(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k dr \xrightarrow{p} \sqrt{T} \int_0^1 f_{\delta_T}(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k dr. \quad (3.C.8)$$

Next, by the telescope trick:

$$\begin{aligned} & \sqrt{T} \int_0^1 f_{\delta_T}(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k(r) dr \\ &= \sqrt{T} \sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \int_0^1 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\} B_{x,i,T}^k(r) dr \\ & \quad - \sqrt{T} \sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \int_0^1 1\{0 \leq \sqrt{T}B_{z,i,T}(r) < \delta_T\} B_{x,i,T}^k(r) dr \\ & \quad + \sqrt{T} \sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \int_0^1 1\{0 \leq \sqrt{T}B_{z,i,T}(r) < \delta_T\} B_{x,i,T}^k(r) dr, \\ &= \sqrt{T} \left(\sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \right) \int_0^1 1\{0 \leq \sqrt{T}B_{z,i,T}(r) < \delta_T\} B_{x,i,T}^k(r) dr + \bar{R}_{i,T}, \end{aligned} \quad (3.C.9)$$

where

$$\bar{R}_{i,T} = \sqrt{T} \sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \int_0^1 [1_{k\delta} - 1_\delta] B_{x,i,T}^k(r) dr,$$

with

$$1_{k\delta} = 1\{k\delta_T \leq \sqrt{T}B_{z,i,T}(r) < (k+1)\delta_T\},$$

and

$$1_\delta = 1\{0 \leq \sqrt{T}B_{z,i,T}(r) < \delta_T\}.$$

Because $f(\cdot)$ is integrable on \mathbb{R} , $\sup_x f(x)$ is bounded. Then, by the triangular inequality:

$$\begin{aligned} |\bar{R}_{i,T}| &= \left| \sqrt{T} \sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \int_0^1 [1_{k\delta} - 1_\delta] B_{x,i,T}^k(r) dr \right|, \\ &\leq \sqrt{T} \sum_{k=-\kappa_T}^{\kappa_T-1} |f(k\delta_T)| \int_0^1 |1_{k\delta} - 1_\delta| B_{x,i,T}^k(r) dr, \\ &\leq \sqrt{T} \sup_k |f(k\delta_T)| \sum_{k=-\kappa_T}^{\kappa_T-1} \int_0^1 |1_{k\delta} - 1_\delta| B_{x,i,T}^k(r) dr, \\ &= \sqrt{T} \sup_k |f(k\delta_T)| \sum_{k=-\kappa_T}^{\kappa_T-1} \int_0^1 B_{x,i,T}^k(r) d(N_{z,i,T}(k\delta_T, (k+1)\delta_T) - N_{z,i,T}(0, \delta_T)), \\ &= O_p(\kappa_T \sqrt{\frac{\delta_T}{\sqrt{T}}}) = o(1). \end{aligned}$$

The last equality is due to the following argument based on Lemma 2.5(a) of Park and Phillips (1999),

$$N_{z,i,T}(k\delta_T, (k+1)\delta_T) - N_{z,i,T}(0, \delta_T) = O_p\left(\sqrt{\frac{\delta_T}{T^{3/2}}}\right),$$

and following (3.C.3),

$$\kappa_T \sqrt{\frac{\delta_T}{\sqrt{T}}} \rightarrow 0,$$

as $T \rightarrow \infty$. Moreover, define $N_{z,i,T}(0, \delta_T; r) = \int_0^r 1\{0 \leq \sqrt{T}B_{z,i,T}(r) < \delta_T\}dr$ for any $0 \leq r \leq 1$,

$$\begin{aligned} & \sqrt{T} \left(\sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \right) \int_0^1 1\{0 \leq \sqrt{T}B_{z,i,T}(r) < \delta_T\} B_{x,i,T}^k(r) dr, \\ &= \sqrt{T} \left(\sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \right) \int_0^1 B_{x,i,T}^k(r) dN_{z,i,T}(0, \delta_T; r), \\ &= \left(\sum_{k=-\kappa_T}^{\kappa_T-1} f(k\delta_T) \delta_T \right) \int_0^1 B_{x,i,T}^k(r) d\frac{\sqrt{T}}{\delta_T} N_{z,i,T}(0, \delta_T; r), \\ &= \left(\int_{-\infty}^{\infty} f_{\delta_T}(s) ds \right) \int_0^1 B_{x,i,T}^k(r) d\frac{\sqrt{T}}{\delta_T} N_{z,i,T}(0, \delta_T; r), \\ &= \left(\int_{-\infty}^{\infty} f(s) ds \right) \int_0^1 B_{x,i,T}^k(r) dL_{z,i}(r, 0) + o_p(1). \end{aligned} \tag{3.C.10}$$

The convergence follows from the inequalities (3.C.1), (3.C.2) and Lemma 2.5(b) of Park and Phillips (1999):

$$\frac{\sqrt{T}}{\delta_T} N_{z,i,T}(0, \delta_T; r) \xrightarrow{p} L_{z,i}(r, 0).$$

Moreover,

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\delta_T}(s) ds &= \int_{-\infty}^{\infty} f_T(s) ds + o(\kappa_T \delta_T^2), \\ \int_{-\infty}^{\infty} f_T(s) ds &= \int_{-\infty}^{\infty} f(s) ds + o(1). \end{aligned}$$

Hence, we have (3.C.5) from (3.C.8), (3.C.9) and (3.C.10). Next, we show (3.C.6) and (3.C.7). Taking T sufficiently large, $f'_T(x)$ and $f''_T(x)$ can be assumed to be monotone as $|x| \rightarrow \infty$, given f is integrable. Let

$\epsilon_T = \sup_{0 \leq r \leq 1} |B_{v,i,T}(r) - B_{v,i}(r)|$, we have the following by Cauchy-Schwarz inequality:

$$\begin{aligned}
& \sqrt{T} \int_0^1 f'_T(\sqrt{T}B_{z,i,T}(r))B_{x,i,T}^k dr \\
& \leq \sqrt{T} \sqrt{\int_0^1 f^2(\sqrt{T}(B_{z,i}(r) - \epsilon_T))1\{\sqrt{T}(B_{z,i}(r) - \epsilon_T) > \kappa_T \delta_T\} dr \int_0^1 (B_{x,i}(r) - \epsilon_T)^{2k} dr}, \\
& = \sqrt{\sqrt{T}\sqrt{T} \int_{-\infty}^{\infty} f^2(\sqrt{T}(s - \epsilon_T))1\{\sqrt{T}(s - \epsilon_T) > \kappa_T \delta_T\} L_{z,i}(1, s) ds \int_0^1 (B_{x,i}(r) - \epsilon_T)^{2k} dr}, \\
& = \sqrt{\int_{-\infty}^{\infty} f^2(x)1\{x > \kappa_T \delta_T - 2\sqrt{T}\epsilon_T\} L_{z,i}(1, x/\sqrt{T} - \epsilon_T) \sqrt{T} dx \int_0^1 (B_{x,i}(r) - \epsilon_T)^{2k} dr}, \\
& = \sqrt{\int_{-\infty}^{\infty} f^2\left(\frac{y}{\sqrt{T}}\right)1\left\{\frac{y}{\sqrt{T}} > \kappa_T \delta_T - 2\sqrt{T}\epsilon_T\right\} L_{z,i}\left(1, \frac{y}{T} - \epsilon_T\right) dy \int_0^1 (B_{x,i}(r) - \epsilon_T)^{2k} dr}, \\
& \xrightarrow{p} 0.
\end{aligned} \tag{3.C.11}$$

The first equality follows from the Local Time lemma (Lemma 2.4 in Park and Phillips (1999)). Similarly, $\int_0^1 (B_{x,i}(r) - \epsilon_T)^{2k} dr$ could be proved to be $O_p(1)$. The last two equalities follow from $x = \sqrt{T}(s - \epsilon_T)$ and $y = \sqrt{T}x$. The convergence is due to $\kappa_T \delta_T - 2\sqrt{T}\epsilon_T \xrightarrow{p} \infty$. Similarly, for f_T'' , (3.C.7) is proved to be $o_p(1)$. Therefore, using (3.C.5), (3.C.6) and (3.C.7), we have

$$\frac{1}{(\sqrt{T})^{(1+k)}} \sum_{t=1}^T x_{i,t}^k f(z_{i,t}) \xrightarrow{p} \int_{-\infty}^{\infty} f(s) ds \int_0^1 B_{x,i}^k(r) dL_{z,i}(r, 0). \tag{3.C.12}$$

Having established (3.C.12), we can use it to prove the sample covariance asymptotics now. First, a continuous martingale process is defined as

$$\begin{aligned}
M_{i,T}(r) = & \sqrt[4]{T} \sum_{t=1}^{l-1} f\left(\sqrt{T}B_{z,i,T}\left(\frac{t-1}{T}\right)\right) B_{x,i,T}^k\left(\frac{t-1}{T}\right) \left(B_{u,i}\left(\frac{\tau_{i,T,t}}{T}\right) - B_{u,i}\left(\frac{\tau_{i,T,t-1}}{T}\right)\right) \\
& + \sqrt[4]{T} f\left(\sqrt{T}B_{z,i,T}\left(\frac{l-1}{T}\right)\right) B_{x,i,T}^k\left(\frac{t-1}{T}\right) \left(B_{u,i}(r) - B_{u,i}\left(\frac{\tau_{i,T,l-1}}{T}\right)\right),
\end{aligned}$$

for $\tau_{i,T,l-1}/T < r \leq \tau_{i,T,l}/T$, where $\tau_{i,T,l}, l = 1, \dots, T$, are the stopping times introduced in Lemma 2.1 of Park and Phillips (2001). Then,

$$\frac{1}{(\sqrt{T})^{1/2+k}} \sum_{t=1}^T f(z_{i,t}) x_{i,t}^k u_{i,t} =_d M_{i,T}\left(\frac{\tau_{i,TT}}{T}\right), \tag{3.C.13}$$

and

$$\sup_{1 \leq t \leq T} \left| \left(\frac{\tau_{i,T,t}}{T} - \frac{\tau_{i,T,t-1}}{T} \right) - \frac{1}{T} \right| \xrightarrow{a.s.} 0 \tag{3.C.14}$$

resulting from Lemma 2.1 of Park and Phillips (2001). The quadratic variation process $[M_{i,T}]$ of $M_{i,T}$ is given by

$$\begin{aligned} [M_{i,T}]_r &= \sqrt{T} \sum_{t=1}^{l-1} f^2 \left(\sqrt{T} B_{z,i,T} \left(\frac{t-1}{T} \right) \right) B_{x,i,T}^{2k} \left(\frac{t-1}{T} \right) \left(\frac{\tau_{i,T,t}}{T} - \frac{\tau_{i,T,t-1}}{T} \right) \\ &\quad + \sqrt{T} f^2 \left(\sqrt{T} B_{z,i,T} \left(\frac{l-1}{T} \right) \right) B_{x,i,T}^{2k} \left(\frac{t-1}{T} \right) \left(r - \frac{\tau_{i,T,l-1}}{T} \right), \\ &= \sqrt{T} \int_0^r f^2 \left(\sqrt{T} B_{z,i,T}(\bar{r}) \right) B_{x,i,T}^{2k}(\bar{r}) d\bar{r}, \\ &\xrightarrow{p} \left(\int_{-\infty}^{\infty} f^2(s) ds \right) \int_0^r B_{x,i,T}^{2k}(\bar{r}) dL(\bar{r}, 0). \end{aligned}$$

The convergence in probability follows from (3.C.12). $f(\cdot)$ is square integrable defined by Assumption I, hence, $f^2(\cdot)$ is integrable. Moreover, the covariation process of $M_{i,T}$ and $B_{v,i}$ is given by

$$\begin{aligned} [M_{i,T}, B_{v,i}]_r &= \sqrt{T} \sum_{t=1}^{l-1} f \left(\sqrt{T} B_{z,i,T} \left(\frac{t-1}{T} \right) \right) B_{x,i,T}^{2k} \left(\frac{t-1}{T} \right) \left(\frac{\tau_{i,T,t}}{T} - \frac{\tau_{i,T,t-1}}{T} \right) \omega_{vu,i} \\ &\quad + \sqrt{T} f \left(\sqrt{T} B_{z,i,T} \left(\frac{l-1}{T} \right) \right) B_{x,i,T}^{2k} \left(\frac{t-1}{T} \right) \left(r - \frac{\tau_{i,T,l-1}}{T} \right) \omega_{vu,i}, \\ &\xrightarrow{p} 0. \end{aligned}$$

The convergence here results from the Modulus Inequality. Park and Phillips (2001) pointed out that Theorem 2.3 (Page 496) in Revuz and Yor (1994) shows that the asymptotic distribution of $M_{i,T}(r)$ is completely determined by $[M_{i,T}]_r$ and $[M_{i,T}, B_{v,i}]_{\rho_{i,T}(r)}$ where $\rho_{i,T}(r) = \inf\{s \in [0, 1] : [M_{i,T}]_s > r\}$. Defining the Dambis, Dubins-Schwarz Brownian motion of $M_{i,T}$ as $W_{i,T}(r) = M_{i,T}(\rho_{i,T}(r))$, $(B_{v,i}, W_{i,T})$ converges jointly in distribution to two independent Brownian motions $(B_{v,i}, W)$ (See Revuz and Yor (1994), Theorem 1.6, Page 173). By the Dambis, Dubins-Schwarz time-change theorem, any continuous martingale $M_{i,T}(r)$ with its quadratic variation $[M_{i,T}]_{\infty} = \infty$ a.s. can be written as $M_{i,T}(r) = W_{[M_{i,T}]_{\rho}}$, where W_i is a Brownian motion defined in the same probability space of $M_{i,T}(r)$, and $[M_{i,T}]_{\rho}$ is defined at the stopping time of $\rho_{i,T}(r)$. Then, the Brownian motion W_i is named Dambis, Dubins-Schwarz Brownian motion of $M_{i,T}(r)$. Hence, the lemma follows from (3.C.13) and

$$\begin{aligned} M_{i,T} \left(\frac{\tau_{i,T,T}}{T} \right) &= M_{i,T}(1) + o_p(1), \\ &\xrightarrow{d} W_i \left(\int_{-\infty}^{\infty} f^2(s) ds \int_0^1 B_{x,i}^{2k}(r) dL_{z,i}(r, 0) \right). \end{aligned}$$

□

3.C.3 Lemma 3.C.2

Lemma 3.C.2. *Let Assumptions 3.1', 3.2 and 3.4 hold with $c = \theta$ and $C = \Theta$, and let $f(x_{i,t}): \mathbb{R} \rightarrow \mathbb{R}$ with $\sum_{t=1}^T f^2(x_{i,t})$ being $S(d)$. Then, for any $\delta > 0$,*

$$\frac{1}{T^{(1+d)/2+\delta}} \sum_{t=1}^T [1\{z_{i,t} > c\}f(x_{i,t}) - 1\{z_{i,t} > 0\}f(x_{i,t})] = o_p(1),$$

uniformly over C .

3.C.4 Proof of Lemma 3.C.2

Without loss of generality, we assume that $C \subset \mathbb{R}_+$. Then, for any arbitrary $c \in C$,

$$\begin{aligned} & \sum_{t=1}^T [1\{z_{i,t} > c\}f(x_{i,t}) - 1\{z_{i,t} > 0\}f(x_{i,t})] \\ & \leq \sqrt{\sum_{t=1}^T [1\{z_{i,t} > c\} - 1\{z_{i,t} > 0\}]^2 \sum_{t=1}^T f(x_{i,t})^2}, \\ & = \sqrt{\sum_{t=1}^T 1\{c > z_{i,t} > 0\} \sum_{t=1}^T f(x_{i,t})^2}, \\ & = \sqrt{T \int_0^1 1\{\frac{c}{\sqrt{T}} > \frac{z_{i,t}(Tr)}{\sqrt{T}} > 0\} dr \sum_{t=1}^T f(x_{i,t})^2}, \\ & =^d \sqrt{T \int_0^1 1\{\frac{c}{\sqrt{T}} > B_{z,i,T} > 0\} dr \sum_{t=1}^T f(x_{i,t})^2}, \\ & = \sqrt{c\sqrt{T} \frac{\sqrt{T}}{c} \int_0^1 1\{\frac{c}{\sqrt{T}} > B_{x,i,T} > 0\} \sum_{t=1}^T f(x_{i,t})^2}, \\ & = \sqrt{c\sqrt{T} [L_{z,i}(1, 0) + o_p(1)] \sum_{t=1}^T f(x_{i,t})^2}, \\ & = O_p(T^{(1+d)/2}), \end{aligned}$$

where $B_{z,i,T}$ is an element of $B_{v,i,T}$. The convergence to the local time follows from Lemma 2.5 (b) from Park and Phillips (1999). Here, we take

$$1/2 < 2/3 - 1/(3p) - \varepsilon,$$

and

$$\varepsilon < 1/6 - 1/(3p) < 1/8,$$

where $\varepsilon > 0$ is defined by Lemma 2.5 (b) from Park and Phillips (1999), and $p > 8$ is defined by Assumption 3.1 (f). □

3.C.5 Proof of Lemma 3.1

1, First part(convergence in probability):

We need to verify the convergence in probability uniformly in Θ . According to Theorem 3.1 of Pötscher and Prucha (1994), pointwise convergence and uniform equicontinuity on a compact set imply uniform convergence. Hence, we need to verify uniform equicontinuity on Θ . Given Θ is compact, we only need to verify equicontinuity at each point. Then, according to Lemma A3 and I -regularity condition (b) of Park and Phillips (2001),

$$\sup_{\theta \in \Theta} f(z_{i,t}, \theta),$$

and

$$\inf_{\theta \in \Theta} f(z_{i,t}, \theta),$$

are both I -regular. Hence, according to Lemma 3.C.1,

$$\frac{1}{(\sqrt{T})^{(1+k)}} \sum_{t=1}^T \sup_{\theta \in \Theta} f(z_{i,t}, \theta)(x_{i,t})^k \xrightarrow{P} \int_{-\infty}^{\infty} \sup_{\theta \in \Theta} f(s, \theta) ds \int_0^1 B_{x,i}^k(r) dL_{z,i}(r, 0), \quad (3.C.15)$$

and,

$$\frac{1}{(\sqrt{T})^{(1+k)}} \sum_{t=1}^T \inf_{\theta \in \Theta} f(z_{i,t}, \theta)(x_{i,t})^k \xrightarrow{P} \int_{-\infty}^{\infty} \inf_{\theta \in \Theta} f(s, \theta) ds \int_0^1 B_{x,i}^k(r) dL_{z,i}(r, 0). \quad (3.C.16)$$

Let N_{δ_0} be a neighbourhood defined for an arbitrary $\theta^0 \in \Theta$. According to the I -regularity condition (a)(equicontinuity) from Park and Phillips (2001),

$$\sup_{\theta \in N_{\delta_0}} f(s, \theta) - \inf_{\theta \in N_{\delta_0}} f(s, \theta) \rightarrow 0,$$

as $\delta_0 \rightarrow 0$. Hence,

$$\int_{-\infty}^{\infty} \sup_{\theta \in \Theta} f(s, \theta) ds - \int_{-\infty}^{\infty} \inf_{\theta \in \Theta} f(s, \theta) ds \rightarrow 0, \quad (3.C.17)$$

as $\delta_0 \rightarrow 0$. (3.C.17) implies (3.C.15)–(3.C.16) $\rightarrow 0$. Hence, according to the conventional 3ϵ method,

$$\frac{1}{(\sqrt{T})^{(1+k)}} \sum_{t=1}^T f(z_{i,t}, \theta)(x_{i,t})^k \xrightarrow{P} \int_{-\infty}^{\infty} f(s, \theta) ds \int_0^1 B_{x,i}^k(r) dL_{z,i}(r, 0),$$

uniformly in N_{θ_0} . Given that θ^0 is arbitrary and Θ is compact, the uniform convergence in Θ holds.

2, Second part(convergence in distribution):

Given the uniform convergence result from the first part, the proof is done with a modification of the second

part of the proof of Lemma 3.C.1. We need to replace the convergence of quadratic variation process $[M_{i,T}]_r$ and the covariance process $[M_{i,T}, B_{v,i}]_{\rho_{i,T}(r)}$ by the uniform convergence proved in the first part. \square

3.C.6 Proof of Lemma 3.2

Follows Lemma 2 from Jennrich (1969) and Remark 3.3 (1). \square

3.C.7 Proof of Lemma 3.3

Given that $\hat{\theta}_{n,T}$ is the minimizer of $Q_{n,T}(\theta)$, we have

$$\begin{aligned}
0 &\geq R_{n,T}^*(\hat{\theta}_{n,T}, \theta_0) \\
&= Q_{n,T}(\hat{\theta}_{n,T}) - Q_{n,T}(\theta_0), \\
&= \sum_{i=1}^n \sum_{t=1}^T [y_{i,t} - g(z_{i,t}, x_{i,t}, \hat{\theta}_{n,T})]^2 - \sum_{i=1}^n \sum_{t=1}^T [y_{i,t} - g(z_{i,t}, x_{i,t}, \theta_0)]^2, \\
&= \sum_{i=1}^n \sum_{t=1}^T [g(z_{i,t}, x_{i,t}, \theta_0) + u_{i,t} - g(z_{i,t}, x_{i,t}, \hat{\theta}_{n,T})]^2 - \sum_{i=1}^n \sum_{t=1}^T u_{i,t}^2, \\
&= \sum_{i=1}^n \sum_{t=1}^T [g(z_{i,t}, x_{i,t}, \theta_0) - g(z_{i,t}, x_{i,t}, \hat{\theta}_{n,T})]^2, \\
&\quad - 2 \sum_{i=1}^n \sum_{t=1}^T [g(z_{i,t}, x_{i,t}, \theta_0) - g(z_{i,t}, x_{i,t}, \hat{\theta}_{n,T})] u_{i,t}, \\
&= R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0) - 2R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0).
\end{aligned} \tag{3.C.18}$$

Hence,

$$R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0) \leq R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0),$$

$$M_{n,T}^{-1} R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0) \leq M_{n,T}^{-1} R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0), \tag{3.C.19}$$

for any n and T .

Let $T \rightarrow \infty$ followed by $n \rightarrow \infty$, suppose $\hat{\theta}_{n,T} \xrightarrow{p(a.s.)} \theta_0$ does not hold. Then, we can find a subsequence $\hat{\theta}_{n_k, T_l}$ of $\hat{\theta}_{n,T}$ that converges to $\tilde{\theta} \neq \theta_0$ because $\hat{\theta}_{n,T}$ is a sequence in a compact set Θ . Due to continuity of $R^*(\theta, \theta_0)$ and uniform convergence,

$$M_{n,T}^{-1} R_{n_k, T_l, 1}^*(\hat{\theta}_{n_k, T_l}, \theta_0) \leq M_{n,T}^{-1} R_{n_k, T_l, 2}^*(\hat{\theta}_{n_k, T_l}, \theta_0) \xrightarrow{p} R^*(\tilde{\theta}, \theta_0) \leq 0,$$

which implies

$$R^*(\tilde{\theta}, \theta_0) = 0, \tag{3.C.20}$$

because $R^*(\theta, \theta_0) \geq 0$ for all θ . (3.C.20) provides a contradiction. Because θ_0 is the unique minimum of $R^*(\theta, \theta_0) \geq 0$,

$$R^*(\tilde{\theta}, \theta_0) = 0,$$

if and only if

$$\tilde{\theta} = \theta_0,$$

which contradicts $\tilde{\theta} \neq \theta_0$. Hence, the assumption of $\hat{\theta}_{n,T}$ not converging in probability (almost surely) to θ_0 is false. \square

3.C.8 Proof of Lemma 3.4

We try to construct a compact and convex set which is sufficiently large in Euclidean (Vector) space to contain θ_0 and $\hat{\theta}_{n,T}$ s. Because $\hat{\theta}_{n,T}$'s are the minimizers of $Q_{n,T}$ for any given n and T , it is necessary that

$$Q_{n,T}(\cdot, \hat{\theta}_{n,T}) - Q_{n,T}(\cdot, \theta_0) < 0.$$

Besides,

$$\lim_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \left[\frac{|Q_{n,T}(\cdot, \hat{\theta}_{n,T}) - Q_{n,T}(\cdot, \theta_0)|}{nM_T} \right] \rightarrow 0.$$

Hence, all values of θ such that

$$Q_{n,T}(\cdot, \theta) - Q_{n,T}(\cdot, \theta_0) > 0,$$

and

$$\lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left[\frac{Q_{n,T}(\cdot, \theta) - Q_{n,T}(\cdot, \theta_0)}{nM_T} \right] > 0,$$

can not be minimizers $\hat{\theta}_{n,T}$. Hence, we need to construct the neighbourhood $\delta(\theta_0, M)$ and its boundary with $M > 0$ being the radius such that $\theta \notin \delta(\theta_0, M)$,

$$\lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left[\inf_{\theta} \frac{Q_{n,T}(\cdot, \theta) - Q_{n,T}(\cdot, \theta_0)}{nM_T} \right] > 0 \text{ a.s..}$$

Namely, we need to find a positive real number M such that

$$\lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left[\inf_{\|\theta - \theta_0\| > M} \left[\frac{Q_{n,T}(\cdot, \theta) - Q_{n,T}(\cdot, \theta_0)}{nM_T} \right] \right] > 0 \text{ a.s..} \quad (3.C.21)$$

Therefore, the compact subset is defined as $\Theta_* = [\theta : \|\theta - \theta_0\| \leq M, \text{ for } M > 0]$ of the parameter space with the constant M being such that (3.C.21) holds. By the definition of $Q_{n,T}$, the limit defined by (3.C.21) is equal to

$$\lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left\{ \inf_{\|\theta - \theta_0\| > M} \left[\frac{\sum^{nT} [y_{i,t} - g(\cdot, \theta)]^2 - \sum^{nT} [y_{i,t} - g(\cdot, \theta_0)]^2}{nM_T} \right] \right\},$$

where $\sum^{nT} = \sum_{i=1}^n \sum_{t=1}^T$,

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left\{ \inf_{\|\theta - \theta_0\| > M} \left[\frac{\sum^{nT} [g(\cdot, \theta_0) - g(\cdot, \theta)]^2 - 2 \sum^{nT} [g(\cdot, \theta_0) - g(\cdot, \theta)] u_{i,t}}{nM_T} \right] \right\}, \\ &\geq \lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left\{ \inf_{\|\theta - \theta_0\| > M} \left[\frac{\sum^{nT} [g(\cdot, \theta_0) - g(\cdot, \theta)]^2 - 2 \sum^{nT} [g(\cdot, \theta_0) - g(\cdot, \theta)] u_{i,t}}{nM_T} \right] \right\}. \end{aligned}$$

By the Triangular Inequality, we have the following,

$$\geq \lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left\{ \inf_{\|\theta - \theta_0\| > M} \left[\frac{\sum^{nT} [g(\cdot, \theta_0) - g(\cdot, \theta)]^2 - 2 \sum^{nT} [|g(\cdot, \theta_0) - g(\cdot, \theta)| |u_{i,t}|]}{nM_T} \right] \right\},$$

by the Cauchy Schwartz Inequality,

$$\geq \lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left\{ \inf_{\|\theta - \theta_0\| > M} \left\{ \sqrt{\frac{\sum^{nT} C_{i,t}^2}{nM_T}} \left[\sqrt{\frac{\sum^{nT} C_{i,t}^2}{nM_T}} - 2 \sqrt{\frac{\sum^{nT} u_{i,t}^2}{nT}} \right] \right\} \right\},$$

where $C_{i,t} = g(\cdot, \theta) - g(\cdot, \theta_0)$. Because $M_T \geq T$,

$$\begin{aligned} &\geq \lim_{n \rightarrow \infty} \inf_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \inf_{T \rightarrow \infty} \left\{ \inf_{\|\theta - \theta_0\| > M} \left\{ \sqrt{\frac{\sum^{nT} C_{i,t}^2}{nM_T}} \left[\sqrt{\frac{\sum^{nT} C_{i,t}^2}{nM_T}} - 2 \sqrt{\frac{\sum^{nT} u_{i,t}^2}{nT}} \right] \right\} \right\}, \\ &\geq MC^{1/2}(MC^{1/2} - 2\sigma) > 0 \text{ a.s.}, \end{aligned}$$

where $C = \mathbf{E} \left[F \left(\inf_{\Theta} \left\| \frac{\partial g(\cdot, \theta)}{\partial \theta} \right\|^2 \right) \cos^2(\gamma) \right]$ and

$$F \left(\inf_{\Theta} \left\| \frac{\partial g(\cdot, \theta)}{\partial \theta} \right\|^2 \right) = \lim_{M_T \rightarrow \infty} \inf_{M_T \rightarrow \infty} \sum_{t=1}^T \inf_{\Theta} \left\| \frac{\partial g(\cdot, \theta)}{\partial \theta} \right\|^2.$$

The last inequality is proved as follows. According to the vector form of the mean value theorem for $C_{i,t} = g(\cdot, \theta) - g(\cdot, \theta_0)$, we have

$$C_{i,t} = \left\langle \frac{\partial g(\cdot, \bar{\theta})}{\partial \theta'} \cdot (\theta - \theta_0) \right\rangle,$$

where $\langle \cdot \rangle$ stands for the inner product. Then,

$$\left\langle \frac{\partial g(\cdot, \bar{\theta})}{\partial \theta'} \cdot (\theta - \theta_0) \right\rangle = \left\| \frac{\partial g(\cdot, \bar{\theta})}{\partial \theta'} \right\| \|\theta - \theta_0\| \cos(\gamma),$$

where γ is the angle between the tangent space and the parameter space, and $\bar{\theta}$ is the vector lying on the segment of $\theta - \theta_0$. Now, define

$$c(\theta) = \inf_{\Theta} \left\| \frac{\partial g(\cdot, \theta)}{\partial \theta} \cos(\gamma) \right\|^2.$$

Then,

$$C_{i,t} \geq M\sqrt{c(\theta)},$$

for $\|\theta - \theta_0\| > M$. Because, $g(\cdot, \theta)$ is twice continuous differentiable, $\cos(\gamma) \neq 0$. Therefore, $\mathbf{E} \left[F \left(\inf_{\Theta} \left\| \frac{\partial g(\cdot, \theta)}{\partial \theta} \right\|^2 \right) \right]$ implies $C > 0$. Moreover,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \liminf_{T \rightarrow \infty} \left\{ \inf_{\|\theta - \theta_0\| > M} \left[\sqrt{\frac{\sum^{nT} C_{i,t}^2}{nM_T}} - 2\sqrt{\frac{\sum^{nT} u_{i,t}^2}{nT}} \right] \right\} &\geq \liminf_{n \rightarrow \infty} \liminf_{T \rightarrow \infty} \left\{ M\sqrt{\frac{\sum^{nT} c(\theta)}{nM_T}} - 2\sqrt{\frac{\sum^{nT} u_{i,t}^2}{nT}} \right\}, \\ &= MC^{1/2} - 2\sigma > 0 \text{ a.s.}, \end{aligned}$$

if $M > 2\sigma/C^{1/2}$. The last equality is due to the continuity of the function $\sqrt{\cdot}$. Hence, we can move the limit sign inside. Given $C > 0$, $M > 0$ and the desired neighbourhood $\delta(\theta, M)$ with its boundary is constructed. Assumption 3.4 implies that there is a compact set of the parameter space, which is defined by M such that all values of parameters of interest are in the set. \square

3.C.9 Proof of Theorem 3.1

We need to verify the sufficient conditions in Lemma 3.3. Condition (c) follows from Assumption 3.4. For condition (b),

- As shown later, $R^*(\theta, \theta_0)$ is the expectation of nonnegative random variables due to the n dimension asymptotics. Hence, it is deterministic and nonnegative.
- Continuity follows from uniform convergence. Because $Q_{n,T}$ is continuous, $R_{n,T}^*$ is the difference of two continuous functions and therefore it is continuous. Uniform convergence keeps the continuity property.
- The uniqueness of θ_0 is due to $\int_{-\infty}^{\infty} [f(s, \theta) - f(s, \theta_0)]^2 ds > 0$ for all $\theta \neq \theta_0$.

Now, we only need to define $M_{n,T}$, $R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0)$ and $R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0)$, and verify the following uniform convergence

$$M_{n,T}^{-1} R_{n,T,1}^*(\theta, \theta_0) \xrightarrow{P} R^*(\theta, \theta_0), \quad (3.C.22)$$

$$M_{n,T}^{-1} R_{n,T,2}^*(\theta, \theta_0) \xrightarrow{P} 0. \quad (3.C.23)$$

According to (3.7) in Section 3.2 and (3.10) in Section 3.3, we have

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0))^2,$$

and,

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0)) u_{i,t},$$

where $g(z_{i,t}, x_{i,t}, \theta) = f(z_{i,t}, \theta) x_{i,t}$.

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))^2 x_{i,t}^2,$$

and,

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0)) x_{i,t} u_{i,t}.$$

Following Lemma A6 of Park and Phillips (2001), $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))^2$ and $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))$ are I -regular. Hence, according to Lemma 3.1, we take $k = 2$ and $M_{n,T} = nT^{3/2}$ for $R_{n,T,1}^*(\theta, \theta_0)$:

$$\frac{R_{n,T,1}^*(\theta, \theta_0)}{nT^{3/2}} \xrightarrow{p} \frac{\sum_{i=1}^n \left(\int_{-\infty}^{\infty} [f(s, \theta) - f(s, \theta_0)]^2 ds \right) \left(\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right)}{n},$$

uniformly in Θ as $T \rightarrow \infty$. Because of the assumption of IID across sections, we have the convergence result in (3.C.22) as $n \rightarrow \infty$. Moreover, based on Lemma 3.1, we have

$$R_{n,T,2}^*(\theta, \theta_0) = nO_p(T^{3/4}),$$

as $T \rightarrow \infty$ for fixed n . Hence,

$$\frac{R_{n,T,2}^*(\theta, \theta_0)}{nT^{3/2}} = \frac{1}{T^{3/2}} O_p(T^{3/4}) \rightarrow 0,$$

uniformly in Θ because $\frac{1}{T^{3/4}} \rightarrow 0$ independent from θ . Finally, (3.C.23) holds. So far, the above convergence results are only pointwise in Θ as $n \rightarrow \infty$. Just applying the same arguments as those in the proof of Lemma 3.1, we can prove that this pointwise convergence implies uniform convergence in Θ given its compactness and the Lipschitz condition from Definition 3.3 of Park and Phillips (2001). \square

3.C.10 Proof of Theorem 3.2

The arguments are similar to those in the proof of Theorem 3.1. The difference is due to $M_{n,T}$, $R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0)$ and $R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0)$. Following (3.7) in Section 3.2 and (3.10) in Section 3.3, we have

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))^2 x_{i,t}^2,$$

and,

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T (f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0)) x_{i,t} u_{i,t}.$$

According to Lemma A6 of Park and Phillips (2001), $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))^2$ and $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))$ are H -regular with asymptotic order $(\nu(\sqrt{T}))^2$ and $\nu(\sqrt{T})$ respectively. We take $M_{n,T} = n(T\nu(\sqrt{T}))^2$, by Theorem 3.3 of Park and Phillips (2001),

$$\frac{R_{n,T,1}^*(\theta, \theta_0)}{n(T\nu(\sqrt{T}))^2} \xrightarrow{d} \frac{\sum_{i=1}^n \int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 B_{x,i}^2(r) dr}{n},$$

uniformly in Θ as $T \rightarrow \infty$. To see the lower bound, we apply the Reverse Hölder inequality, because $\inf_{r \in [0,1]} B_{x,i}^2(r) \neq 0$ almost surely.

$$\begin{aligned} \int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 B_{x,i}^2(r) dr &\geq \sqrt{\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 dr} \left[\int_0^1 B_{x,i}^2(r) dr \right]^{-1}, \\ &= \sqrt{\int_{-\infty}^{\infty} [h(s, \theta) - h(s, \theta_0)]^2 L_{z,i}(1, s) ds} \left[\int_0^1 B_{x,i}^2(r) dr \right]^{-1}. \end{aligned}$$

The last equality is due to Lemma 2.4 of Park and Phillips (1999). Because of the assumption of IID across sections, we have the following for $n \rightarrow \infty$,

$$\frac{R_{n,T,1}^*(\theta, \theta_0)}{n(T\nu(\sqrt{T}))^2} \xrightarrow{a.s.} \mathbf{E} \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 B_{x,i}^2(r) dr \right).$$

Moreover, based on Theorem 3.3 of Park and Phillips (2001), we have

$$R_{n,T,2}^*(\theta, \theta_0) = nO_p(T\nu(\sqrt{T})),$$

as $T \rightarrow \infty$ for fixed n . Hence,

$$\frac{R_{n,T,2}^*(\theta, \theta_0)}{n(T\nu(\sqrt{T}))^2} = \frac{1}{T\nu(\sqrt{T})} O_p(T\nu(\sqrt{T})) \rightarrow 0,$$

uniformly in Θ because $\frac{1}{T\nu(\sqrt{T})} \rightarrow 0$ independent from θ . The uniform convergence follows from the same arguments in the proof of Theorem 3.1. However, the Lipschitz condition is replaced by the locally equicontinuity condition introduced in Definition 3.2 of Park and Phillips (2001). \square

3.C.11 Proof of Lemma 3.5

See Park and Phillips (2001). \square

3.C.12 Proof of Theorem 3.3

According to Lemma 3.5, we only need to verify conditions (a)-(e). Besides, according to Lemma 6(b) from Park and Phillips (2001), all products of two I -regular functions are still I -regular, such as $f^2(z_{i,t}, \theta_0)$ or $f(z_{i,t}, \theta_0)\dot{f}(z_{i,t}, \theta_0)$. Hence, we can apply Lemma 3.1 directly.

Condition (a): From the definition of $g(z_{i,t}, x_{i,t}, \theta)$ and $Q_{n,T}(\theta)$, we have

$$\dot{g}(z_{i,t}, x_{i,t}, \theta_0) = \dot{f}(z_{i,t}, \theta_0)x_{i,t}, \quad (3.C.24)$$

$$\dot{Q}_{n,T}(\theta_0) = - \sum_{i=1}^n \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)x_{i,t}u_{i,t}, \quad (3.C.25)$$

and,

$$\begin{aligned} \ddot{Q}_{n,T}^0(\theta_0) &= \sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta_0)\dot{g}(z_{i,t}, x_{i,t}, \theta_0)', \\ &= \sum_{i=1}^n \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)\dot{f}(z_{i,t}, \theta_0)'x_{i,t}^2. \end{aligned} \quad (3.C.26)$$

Now, let $C_{n,T} = \sqrt{n}(\sqrt{T})^{3/4}$, and we have

$$\begin{aligned} L_{n,T} &= -C_{n,T}^{-1}\dot{Q}_{n,T}(\theta_0), \\ &= \frac{1}{\sqrt{n}(\sqrt{T})^{3/4}} \sum_{i=1}^n \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)x_{i,t}u_{i,t}. \end{aligned}$$

According to Lemma 3.1 and CLT, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$L_{n,T} \xrightarrow{d} \mathbf{N}(0, \Sigma),$$

with

$$\Sigma = \int_{-\infty}^{\infty} \dot{f}(s, \theta_0)\dot{f}(s, \theta_0)'ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right].$$

The above expectation results from the IID property of

$$\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0),$$

across i , which itself results from Assumption 3.1' and 3.2. Next, we establish the convergence of $J_{n,T}$:

$$J_{n,T} = C_{n,T}^{-1}\ddot{Q}_{n,T}^0(\theta_0)C_{n,T}^{-1},$$

$$J_{n,T} = \frac{1}{n(\sqrt{T})^3} \sum_{i=1}^n \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)\dot{f}(z_{i,t}, \theta_0)'x_{i,t}^2.$$

Then, from Lemma 3.1 and conventional LLNs, for $T \rightarrow \infty$ it followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{P} \int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right].$$

Hence, the results of Theorem 3.3 follow from the convergence of

$$C_{n,T}(\hat{\theta}_{n,T} - \theta_0) = C_{n,T} J_{n,T}^{-1} L_{n,T}.$$

Condition (b): According to the definition of $\ddot{G}(z_{i,t}, x_{i,t}, \theta_0)$, we have

$$\ddot{G}(z_{i,t}, x_{i,t}, \theta_0) = \ddot{F}(z_{i,t}, \theta_0) x_{i,t}, \quad (3.C.27)$$

where \ddot{F} is the matrix of second derivatives of f with respect to θ . According to Lemma 3.1, we know that

$$\frac{\sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t}}{(\sqrt{T})^{3/4}} = O_p(1).$$

Hence,

$$\begin{aligned} R_{n,T} &= C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta_0) u_{i,t} C_{n,T}^{-1}, \\ &= o_p(1), \end{aligned} \quad (3.C.28)$$

for $T \rightarrow \infty$ followed by $n \rightarrow \infty$.

Condition(c): For $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{P} J,$$

which follows from the proof of Condition (a). Hence, $J > 0$ results from the assumption $\int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds > 0$. Besides, $\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0)$ is positive because $B_{x,i}^2$ and $L_{z,i}(r, 0)$ are positive.

Condition (d): Given the consistency result from Theorem 3.1 and θ_0 being an interior point, this condition holds trivially.

Condition (e): Let \dot{g} and \ddot{G} defined as (3.C.25) and (3.C.26).

$$\begin{aligned}\ddot{Q}_{n,T}(\theta) &= \sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta) \dot{g}(z_{i,t}, x_{i,t}, \theta)' \\ &\quad + \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta) [g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0)] \\ &\quad - \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta) u_{i,t}.\end{aligned}$$

According to the convergence result of $J_{n,T}$ in Condition (a) and Lemma 3.1, we establish that

$$\frac{\sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta) \dot{g}(z_{i,t}, x_{i,t}, \theta)'}{C_{n,T}^2} \xrightarrow{p} \int_{-\infty}^{\infty} \dot{f}(s, \theta) \dot{f}(s, \theta)' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right], \quad (3.C.29)$$

uniformly in $\theta \in \Theta$. Recall \ddot{g} is the vector form of \ddot{G} . Then, according to (3.C.28) and the proof of Lemma 3.1,

$$\frac{\sum_{i=1}^n \sum_{t=1}^T \ddot{g}(z_{i,t}, x_{i,t}, \theta) u_{i,t}}{C_{n,T}^2} \xrightarrow{p} 0, \quad (3.C.30)$$

uniformly in $\theta \in \Theta$. Because \ddot{f} exist, \ddot{f} is bounded. Besides, according to Park and Phillips (2001), $|f(\cdot, \theta) - f(\cdot, \theta_0)|$ is I -regular.

$$\begin{aligned}& \frac{\|\sum_{i=1}^n \sum_{t=1}^T \ddot{g}(z_{i,t}, x_{i,t}, \theta) [g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0)]\|}{C_{n,T}^2} \\ & \leq \frac{\|B_{\ddot{f}}\| \sum_{i=1}^n \sum_{t=1}^T |f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0)| x_{i,t}^2}{C_{n,T}^2}, \\ & \xrightarrow{p} \|B_{\ddot{f}}\| \mathbf{E} \left(\int_{-\infty}^{\infty} |f(s, \theta) - f(s, \theta_0)| ds \int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right),\end{aligned} \quad (3.C.31)$$

where $\|B_{\ddot{f}}\|$ is the upper bound of \ddot{f} . The convergence follows from Lemma 3.1. If θ converges to θ_0 , (3.C.31) converges to zero. Hence, combining (3.C.29), (3.C.30) and (3.C.30), we have

$$C_{n,T}^{-1}(\ddot{Q}_{n,T}(\theta_{n,T}))C_{n,T}^{-1} \xrightarrow{p} J, \quad (3.C.32)$$

due to the definition of $\theta_{n,T}$ and consistency of $\hat{\theta}_{n,T}$, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$. Besides, according to the proof of Condition (a) and Condition (b)

$$C_{n,T}^{-1}(\ddot{Q}_{n,T}(\theta_0))C_{n,T}^{-1} \xrightarrow{p} J.$$

Hence, Condition (e) is established in view of (3.C.32). \square

3.C.13 Proof of Theorem 3.4

The proof is in the same spirit as that of Theorem 3.3 except for applying different T dimension asymptotics. Besides, by extending the arguments in Lemma 6(c) from Park and Phillips (2001), the product of two or more H -regular functions is still H -regular with the new asymptotic orders being the product of the orders of the factors of the product. Hence, we could apply the nonlinear time series asymptotic results in Park and Phillips (2001) directly. All notations and definitions are the same as those in the proof of Theorem 3.3 except for assuming $f(z_{i,t}, \theta)$ being H -regular.

Condition (a): Now, let $C_{n,T} = \sqrt{n}T\dot{\nu}$. According to Theorem 3.3 from Park and Phillips (2001), we have

$$(T\dot{\nu})^{-1} \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0) x_{i,t} u_{i,t} \xrightarrow{d} \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i} dB_{u,i},$$

as $T \rightarrow \infty$. By applying CLT, for $n \rightarrow \infty$,

$$L_{n,T} \xrightarrow{d} \mathbf{N}(0, \Sigma_H),$$

with

$$\Sigma_H = \mathbf{E} \left[\int_0^1 \dot{h}(B_{z,i}, \theta_0) \dot{h}(B_{z,i}, \theta_0)' B_{x,i}^2 dr \right].$$

The above expectation is follows from the IID property of

$$\int_0^1 \dot{h}(B_{z,i}, \theta_0) \dot{h}(B_{z,i}, \theta_0)' B_{x,i}^2 dr,$$

across i , which results from Assumptions 3.1' and 3.2. Next, we establish the convergence of $J_{n,T}$. According to Theorem 3.3 from Park and Phillips (2001) and LLNs, for $T \rightarrow \infty$ it followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{P} \mathbf{E} \left[\int_0^1 \dot{h}(B_{z,i}, \theta_0) \dot{h}(B_{z,i}, \theta_0)' B_{x,i}^2 dr \right].$$

Hence, the results of Theorem 3.4 follow from the above obtained limits of

$$C_{n,T}(\hat{\theta}_{n,T} - \theta_0) = C_{n,T} J_{n,T}^{-1} L_{n,T}.$$

Condition (b): According to Theorem 3.3 from Park and Phillips (2001),

$$(T\ddot{\nu})^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t} = O_p(1).$$

Hence,

$$\begin{aligned}
& (T\dot{\nu})^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t} \\
&= (\dot{\nu}^{-1} \ddot{\nu})(\ddot{\nu}T)^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t}, \\
&\leq \|\dot{\nu}^{-1} \ddot{\nu}\| (\ddot{\nu}T)^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t}, \\
&= O_p(1).
\end{aligned} \tag{3.C.33}$$

The above inequality follows from the fact that $\dot{\nu}$ and $\ddot{\nu}$ are strictly positive. Furthermore, $\|\dot{\nu}^{-1} \ddot{\nu}\| < \infty$ is due to condition (a) of Assumption H. Hence,

$$\begin{aligned}
R_{n,T} &= C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta_0) u_{i,t} C_{n,T}^{-1}, \\
&= o_p(1),
\end{aligned}$$

for $T \rightarrow \infty$ followed by $n \rightarrow \infty$.

Condition(c): Follows from the same argument in the proof of Theorem 3.3.

Condition (d): Follows from the same argument in the proof of Theorem 3.3.

Condition (e): The arguments are similar to those for the proof of Theorem 3.3 but with the time series asymptotic results of the proof of Theorem 5.2 in Park and Phillips (2001). Hence, we need the following inequalities

$$\begin{aligned}
& (\dot{\nu} \otimes \dot{\nu})^{-1} \\
&= (\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu} \nu^{-1} \nu^{-1}, \\
&\leq \|(\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu}\| \nu^{-1} \nu^{-1}.
\end{aligned}$$

$\ddot{\nu}^{-1} \nu^{-1}$ is the convergence rate of the product of f and \ddot{f} . According to the expression in (3.C.31), they are needed for defining $C_{n,T}$. Because the boundedness of \ddot{f} does not apply to H -regular functions. Furthermore, condition (a) of Assumption H implies the boundedness of $\|(\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu}\|$. \square

3.C.14 Proof of Corollary 3.1

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 3.1, the results in Corollary 3.1 is a natural byproduct of the proof of Theorem 3.3 in view of Condition (e) in Lemma 3.5. \square

3.C.15 Proof of Corollary 3.2

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 3.2, the results in Corollary 3.2 is a natural byproduct of the proof of Theorem 3.4 in view of Condition (e) in Lemma 3.5. \square

3.C.16 Proof of Theorem 3.5

Conditions (b) and (c) in Lemma 3.3 follow from the same arguments as those in the proof of Theorem 3.1. We firstly define $M_{1,n,T}$, $R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0)$ and $R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0)$, and verify the following uniform convergence

$$M_{1,n,T}^{-1} R_{n,T,1}^*(\theta, \theta_0) \xrightarrow{p} R^*(\theta_1, \theta_2 \theta_{1,0}, \theta_{2,0}), \quad (3.C.34)$$

$$M_{1,n,T}^{-1} R_{n,T,2}^*(\theta, \theta_0) \xrightarrow{p} 0. \quad (3.C.35)$$

Let f_1^0 denote $f_1(z_{i,t}; \theta_{1,0}, \theta_{2,0}, \theta_{3,0})$. Similarly, we define f_1^0 , f_2^0 and f_2^0 . We start with the consistency proof of $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})'$. Recall

$$\theta = (\theta'_1, \theta'_2, \theta'_3, \theta'_4)',$$

then, for the Model (3.1) in Section 3.2 and the objective function (3.10) in Section 3.3, we have

$$\begin{aligned} R_{n,T,1}^*(\theta, \theta_0) &= \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (f_1 + f_2)]^2 x_{i,t}^2, \\ R_{n,T,2}^*(\theta, \theta_0) &= \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (f_1 + f_2)] x_{i,t} u_{i,t}. \end{aligned}$$

We take $M_{1,n,T} = n(T\nu_1)^2$ and establish (3.C.35) first. According to Lemma 5(d) from Chang et al. (2001), we have

$$\sum_{t=1}^T (f_1^0 - f_1) x_{i,t} u_{i,t} = O_p(T\nu_1).$$

Hence,

$$M_{1,n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T (f_1^0 - f_1) x_{i,t} u_{i,t} = o_p(1),$$

as $T \rightarrow \infty$ followed by $n \rightarrow \infty$. Similarly, we have

$$\sum_{t=1}^T (f_2^0 - f_2) x_{i,t} u_{i,t} = O_p(T\nu_2).$$

Because $\frac{\nu_2}{\nu_1} = o_p(1)$,

$$M_{1,n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T (f_2^0 - f_2) x_{i,t} u_{i,t} = o_p(1).$$

Therefore, (3.C.35) holds. Now, we try to establish (3.C.34).

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 - f_1)^2 + (f_2^0 - f_2)^2 + 2(f_1^0 - f_1)(f_2^0 - f_2)] x_{i,t}^2.$$

According to Lemma 5 from Chang et al. (2001), we have, as $T \rightarrow \infty$,

$$\frac{1}{(T\nu_1)^2} \sum_{t=1}^T (f_1^0 - f_1)^2 x_{i,t}^2 \xrightarrow{d} \int_0^1 [h_1^0 - h_1]^2 B_{x,i}^2(r) dr,$$

$$\frac{1}{(T\nu_2)^2} \sum_{t=1}^T (f_2^0 - f_2)^2 x_{i,t}^2 = O_p(1),$$

$$\frac{1}{T^2\nu_1\nu_2} \sum_{t=1}^T [2(f_1^0 - f_1)(f_2^0 - f_2)] x_{i,t}^2 = O_p(1).$$

Because $\frac{\nu_2}{\nu_1} = o_p(1)$, combining the above three convergence results and applying the law of large numbers as $n \rightarrow \infty$, we have

$$R^*(\theta, \theta_0) = \mathbf{E} \left[\int_0^1 [h_1^0 - h_1]^2 B_{x,i}^2(r) dr \right],$$

where $h_1^0 = h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})$ and $h_1 = h_1(B_{z,i}; \theta_1, \theta_2)$. The uniform convergence follows from the same arguments as those in the proof of Theorem 3.1 and Theorem 3.2. Applying the argument of Lemma 3.5 on asymptotic normality to $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})'$, we obtain at least $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})' = O_p(T^{-1}\dot{\nu}_1^{-1})$ even if the regressors are not strictly exogenous. Then, we consider the following objective function for $(\theta'_3, \theta'_4)'$ conditional on the $T\dot{\nu}_1$ consistency of $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})'$:

$$R_{n,T,1}^*((\hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3, \theta_4), \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (\hat{f}_1 + \hat{f}_2)]^2 x_{i,t}^2,$$

and,

$$R_{n,T,2}^*((\hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3, \theta_4), \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (\hat{f}_1 + \hat{f}_2)] x_{i,t} u_{i,t}.$$

Let $M_{2,n,T} = n(T\nu_2)^2$, $\hat{f}_1 = f_1(z_{i,t}; \hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3)$, $\hat{f}_2 = f_2(z_{i,t}; \hat{\theta}_{2,n,T}, \theta_3, \theta_4)$, $\bar{f}_1 = f_1(z_{i,t}; \theta_{1,0}, \theta_{2,0}, \theta_3)$ and $\bar{f}_2 = f_2(z_{i,t}; \theta_{2,0}, \theta_3, \theta_4)$. From Lemma 5(d) from Chang et al. (2001) and the Mean Value Theorem, we have,

as $T \rightarrow \infty$,

$$\begin{aligned} \sum_{t=1}^T (f_1^0 - \hat{f}_1) x_{i,t} u_{i,t} &= \sum_{t=1}^T (f_1^0 - \bar{f}_1 + \bar{f}_1 - \hat{f}_1) x_{i,t} u_{i,t}, \\ &\leq \sum_{t=1}^T (f_1^0 - \bar{f}_1) x_{i,t} u_{i,t} + \|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\| \sum_{t=1}^T \|\hat{f}_1\| x_{i,t} u_{i,t}, \\ &= O_p(T\nu_2) + O_p\left(\frac{\dot{\nu}_1}{\nu_1}\right). \end{aligned}$$

Note: As $\|\hat{f}_1\|$ is H -regular, so each element of \hat{f}_1 is H -regular and $(f_1^0 - \bar{f}_1)$ is $\nu_3 = \nu_2$ H -regular with respect to θ_3 .

$$\begin{aligned} \sum_{t=1}^T (f_2^0 - \hat{f}_2) x_{i,t} u_{i,t} &= \sum_{t=1}^T (f_2^0 - \bar{f}_2 + \bar{f}_2 - \hat{f}_2) x_{i,t} u_{i,t}, \\ &\leq \sum_{t=1}^T (f_2^0 - \bar{f}_2) x_{i,t} u_{i,t} + \|\hat{\theta}'_{2,n,T} - \theta'_{2,0}\| \sum_{t=1}^T \|\hat{f}_2\| x_{i,t} u_{i,t}, \\ &= O_p(T\nu_2) + O_p\left(\frac{\dot{\nu}_2}{\nu_1}\right). \end{aligned}$$

Combining the above two results with condition (b), we have

$$M_{2,n,T}^{-1} R_{n,T,2}^* ((\hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3, \theta_4), \theta_0) \xrightarrow{p} 0.$$

Next, considering

$$\begin{aligned} \sum_{t=1}^T (f_1^0 - \hat{f}_1)^2 x_{i,t}^2 &= \sum_{t=1}^T (f_1^0 - \bar{f}_1 + \bar{f}_1 - \hat{f}_1)^2 x_{i,t}^2, \\ &\leq \|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\|^2 \sum_{t=1}^T \|\hat{f}_1\|^2 x_{i,t}^2 + \sum_{t=1}^T (f_1^0 - \bar{f}_1)^2 x_{i,t}^2 \\ &\quad + 2\|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\| \sum_{t=1}^T \|\hat{f}_1\| (f_1^0 - \bar{f}_1) x_{i,t}^2, \\ &= O_p\left(\frac{\dot{\nu}_1^2}{\nu_1^2}\right) + O_p(T^2\nu_2^2) + O_p\left(\frac{\dot{\nu}_1\nu_2T}{\nu_1}\right), \end{aligned}$$

and

$$\begin{aligned} \sum_{t=1}^T (f_2^0 - \hat{f}_2)^2 x_{i,t}^2 &= \sum_{t=1}^T (f_2^0 - \bar{f}_2 + \bar{f}_2 - \hat{f}_2)^2 x_{i,t}^2, \\ &\leq \sum_{t=1}^T (f_2^0 - \bar{f}_2)^2 x_{i,t}^2 + \|\hat{\theta}_{2,n,T} - \theta_{2,0}\|^2 \sum_{t=1}^T \|\hat{f}_2\|^2 x_{i,t}^2 + \|\hat{\theta}_{2,n,T} - \theta_{2,0}\| \sum_{t=1}^T \|\hat{f}_2\| (f_2^0 - \bar{f}_2) x_{i,t}^2, \\ &= O_p(T^2\nu_2^2) + O_p\left(\frac{\dot{\nu}_2^2}{\nu_1^2}\right) + O_p\left(\frac{\nu_2\dot{\nu}_2T}{\nu_1}\right). \end{aligned}$$

Note: $(f_1^0 - \bar{f}_1)$ is H -regular with respect to θ_3 with degree of homogeneity being $\nu_3 = \nu_2$. Similarly, we have

$$\begin{aligned}
& 2 \sum_{t=1}^T (f_2^0 - \hat{f}_2)(f_1^0 - \hat{f}_1)x_{i,t}^2 \\
&= 2 \sum_{t=1}^T (f_2^0 - \bar{f}_2 + \bar{f}_2 - \hat{f}_2)(f_1^0 - \bar{f}_1 + \bar{f}_1 - \hat{f}_1)x_{i,t}^2, \\
&= 2 \sum_{t=1}^T (f_2^0 - \bar{f}_2)(f_1^0 - \bar{f}_1)x_{i,t}^2 + 2 \sum_{t=1}^T (f_2^0 - \bar{f}_2)(\bar{f}_1 - \hat{f}_1)x_{i,t}^2 \\
&\quad + 2 \sum_{t=1}^T (\bar{f}_2 - \hat{f}_2)(f_1^0 - \bar{f}_1)x_{i,t}^2 + 2 \sum_{t=1}^T (\bar{f}_2 - \hat{f}_2)(\bar{f}_1 - \hat{f}_1)x_{i,t}^2, \\
&\leq O_p(T^2\nu_2^2) + 2\|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\| \sum_{t=1}^T \|\hat{f}_1\| (f_2^0 - \bar{f}_2)x_{i,t}^2 \\
&\quad + 2\|\hat{\theta}_{2,n,T} - \theta_{2,0}\| \sum_{t=1}^T \|\hat{f}_2\| (f_1^0 - \bar{f}_1)x_{i,t}^2 \\
&\quad + 2\|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\| \|\hat{\theta}_{2,n,T} - \theta_{2,0}\| \sum_{t=1}^T \|\hat{f}_1\| \|\hat{f}_2\| x_{i,t}^2, \\
&= O_p(T^2\nu_2^2) + O_p\left(\frac{\nu_2\dot{\nu}_1 T}{\nu_1}\right) + O_p\left(\frac{\nu_2\dot{\nu}_2 T}{\nu_1}\right) + O_p\left(\frac{\dot{\nu}_1\dot{\nu}_2}{\nu_1^2}\right).
\end{aligned}$$

Hence, we have

$$M_{2,n,T}^{-1} R_{n,T,1}^* (\hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3, \theta_4, \theta_0) \xrightarrow{p} \mathbf{E} \left(\int_0^1 [\bar{h}_{r_1}^B - h_{r_1}^{B,0} + \bar{h}_2^B - h_2^{B,0}]^2 B_{x,i}^2(r) dr \right),$$

by the limit of

$$M_{2,n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T [(f_2^0 - \bar{f}_2)^2 + (f_1^0 - \bar{f}_1)^2 + 2(f_2^0 - \bar{f}_2)(f_1^0 - \bar{f}_1)] x_{i,t}^2,$$

where $h_1(\cdot; \theta_{1,0}, \theta_{2,0})$ cancels out in both f_1^0 and \bar{f}_1 , and only r_1^0 and \bar{r}_1 remain. Note: If θ_3 appears in h_1 , f_1 is ν_1 H -regular with respect to θ_3 but not ν_3 . \square

3.C.17 Proof of Theorem 3.6

The proof is similar to that of Theorem 3.5. Conditions (b) and (c) in Lemma 3.3 follow from the same arguments as those in the proof of Theorem 3.1. We firstly define $M_{1,n,T}$, $R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0)$ and $R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0)$,

and verify the following uniform convergence

$$M_{1,n,T}^{-1} R_{n,T,1}^*(\theta, \theta_0) \xrightarrow{p} R^*(\theta_1, \theta_2 \theta_{1,0}, \theta_{2,0}), \quad (3.C.36)$$

$$M_{1,n,T}^{-1} R_{n,T,2}^*(\theta, \theta_0) \xrightarrow{p} 0. \quad (3.C.37)$$

For the proof of Theorem 3.6, we start with the consistency proof of $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})'$. According to Model (3.7) in Section 3.2 and (3.10) in Section 3.3, we have

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (f_1 + f_2)]^2 x_{i,t}^2,$$

and,

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (f_1 + f_2)] x_{i,t} u_{i,t}.$$

We take $M_{1,n,T} = n(T\nu_1)^2$ and establish (3.C.35) first. According to Lemma 5(d) from Chang et al. (2001), we have

$$\sum_{t=1}^T (f_1^0 - f_1) x_{i,t} u_{i,t} = O_p(T\nu_1).$$

Hence,

$$M_{1,n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T (f_1^0 - f_1) x_{i,t} u_{i,t} = o_p(1),$$

as $T \rightarrow \infty$ followed by $n \rightarrow \infty$. According to Lemma 3.1, we have

$$\sum_{t=1}^T (f_2^0 - f_2) x_{i,t} u_{i,t} = O_p(T^{3/4}).$$

Hence,

$$M_{1,n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T (f_2^0 - f_2) x_{i,t} u_{i,t} = o_p(1).$$

Therefore, (3.C.37) holds.

Now, we try to establish (3.C.36).

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 - f_1)^2 + (f_2^0 - f_2)^2 + 2(f_1^0 - f_1)(f_2^0 - f_2)] x_{i,t}^2.$$

According to Lemma 5 from Chang et al. (2001), we have, as $T \rightarrow \infty$,

$$\frac{1}{(T\nu_1)^2} \sum_{t=1}^T (f_1^0 - f_1)^2 x_{i,t}^2 \xrightarrow{d} \int_0^1 [h_1^0 - h_1]^2 B_{x,i}^2(r) dr,$$

and

$$\begin{aligned}
& \sum_{t=1}^T [2(f_1^0 - f_1)(f_2^0 - f_2)]x_{i,t}^2 \\
& \leq 2 \sqrt{\sum_{t=1}^T (f_2^0 - f_2)^2 \sum_{t=1}^T (f_1^0 - f_1)^2 x_{i,t}^4} \\
& = O_p(T^{7/4} \nu_1).
\end{aligned}$$

The inequality follows from the Cauchy-Schwarz inequality. According to Lemma 3.1, we have

$$\frac{1}{T^{3/2}} \sum_{t=1}^T (f_2^0 - f_2)^2 x_{i,t}^2 = O_p(1).$$

Note: According to Chang et al. (2001), $(f_2^0 - f_2)^2$ is I -regular. Combining the above three convergence results and applying LLNs as $n \rightarrow \infty$, we have

$$R^*(\theta, \theta_0) = \mathbf{E} \left[\int_0^1 [h_1^0 - h_1]^2 B_{x,i}^2(r) dr \right].$$

The uniform convergence follows using the same arguments as in the proof of Theorem 3.1 and Theorem 3.2. Applying the argument of Lemma 3.5 to asymptotic normality to $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})'$, we obtain at least $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})' = O_p(T^{-1} \nu_1^{-1})$ even if the regressors are not strictly exogenous. Then, we consider the following objective function for $(\theta'_3, \theta'_4)'$ conditional on the $T \nu_1$ consistency of $(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})'$.

$$R_{n,T,1}^*((\hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3, \theta_4), \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (\hat{f}_1 + \hat{f}_2)]^2 x_{i,t}^2,$$

and,

$$R_{n,T,2}^*((\hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3, \theta_4), \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0) - (\hat{f}_1 + \hat{f}_2)] x_{i,t} u_{i,t}.$$

Let $M_{2,n,T} = n(T)^{3/2}$. Then, according to Lemma 5(d) from Chang et al. (2001) and the Mean Value Theorem, we have, as $T \rightarrow \infty$,

$$\begin{aligned}
\sum_{t=1}^T (f_1^0 - \hat{f}_1) x_{i,t} u_{i,t} &= \sum_{t=1}^T (f_1^0 - \bar{f}_1 + \bar{f}_1 - \hat{f}_1) x_{i,t} u_{i,t} \\
&\leq \sum_{t=1}^T (f_1^0 - \bar{f}_1) x_{i,t} u_{i,t} + \|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\| \sum_{t=1}^T \|\hat{f}_1\| x_{i,t} u_{i,t}, \\
&= O_p(T^{3/4}) + O_p\left(\frac{\nu_1}{\nu_1}\right).
\end{aligned}$$

Note: As $\|\hat{f}_1\|$ is H -regular, so each element of \hat{f}_1 is H -regular and $(f_1^0 - \bar{f}_1)$ is I -regular with respect

to θ_3 .

$$\begin{aligned}
\sum_{t=1}^T (f_2^0 - \hat{f}_2) x_{i,t} u_{i,t} &= \sum_{t=1}^T (f_2^0 - \bar{f}_2 + \bar{f}_2 - \hat{f}_2) x_{i,t} u_{i,t} \\
&\leq \sum_{t=1}^T (f_2^0 - \bar{f}_2) x_{i,t} u_{i,t} + \|\hat{\theta}_{2,n,T}' - \theta_{2,0}'\| \sum_{t=1}^T \|\hat{f}_2\| x_{i,t} u_{i,t}, \\
&= O_p(T^{3/4}) + O_p\left(\frac{1}{\nu_1 T^{1/4}}\right).
\end{aligned}$$

Combining the above two results with condition (b), we have

$$M_{2,n,T}^{-1} R_{n,T,2}^* ((\hat{\theta}_{1,n,T}', \hat{\theta}_{2,n,T}', \theta_3, \theta_4), \theta_0) \xrightarrow{p} 0.$$

Next, considering

$$\begin{aligned}
&\sum_{t=1}^T (f_1^0 - \hat{f}_1)^2 x_{i,t}^2 \\
&= \sum_{t=1}^T (f_1^0 - \bar{f}_1 + \bar{f}_1 - \hat{f}_1)^2 x_{i,t}^2, \\
&\leq \|(\hat{\theta}_{1,n,T}', \hat{\theta}_{2,n,T}')' - (\theta_{1,0}', \theta_{2,0}')'\|^2 \sum_{t=1}^T \|\hat{f}_1\|^2 x_{i,t}^2 + \sum_{t=1}^T (f_1^0 - \bar{f}_1)^2 x_{i,t}^2 \\
&\quad + 2\|(\hat{\theta}_{1,n,T}', \hat{\theta}_{2,n,T}')' - (\theta_{1,0}', \theta_{2,0}')'\| \sum_{t=1}^T \|\hat{f}_1\| (f_1^0 - \bar{f}_1) x_{i,t}^2, \\
&\leq O_p\left(\frac{\dot{\nu}_1^2}{\nu_1^2}\right) + O_p(T^{3/2}) + 2\|(\hat{\theta}_{1,n,T}', \hat{\theta}_{2,n,T}')' - (\theta_{1,0}', \theta_{2,0}')'\| \sqrt{\sum_{t=1}^T (f_1^0 - \bar{f}_1)^2 \sum_{t=1}^T \|\hat{f}_1\|^2 x_{i,t}^4}, \\
&= O_p\left(\frac{\dot{\nu}_1^2}{\nu_1^2}\right) + O_p(T^{3/2}) + O_p\left(\frac{\dot{\nu}_1 T^{3/4}}{\nu_1}\right),
\end{aligned}$$

and

$$\begin{aligned}
& \sum_{t=1}^T (f_2^0 - \hat{f}_2)^2 x_{i,t}^2 \\
&= \sum_{t=1}^T (f_2^0 - \bar{f}_2 + \bar{f}_2 - \hat{f}_2)^2 x_{i,t}^2, \\
&\leq \sum_{t=1}^T (f_2^0 - \bar{f}_2)^2 x_{i,t}^2 + \|\hat{\theta}_{2,n,T} - \theta_{2,0}\|^2 \sum_{t=1}^T \|\hat{f}_2\|^2 x_{i,t}^2 \\
&\quad + \|\hat{\theta}_{2,n,T} - \theta_{2,0}\| \sum_{t=1}^T \|\hat{f}_2\| (f_2^0 - \bar{f}_2) x_{i,t}^2, \\
&\leq O_p(T^{3/2}) + O_p\left(\frac{T^{1/2}}{\nu_1^2}\right) + \|\hat{\theta}_{2,n,T} - \theta_{2,0}\| \sqrt{\sum_{t=1}^T (f_2^0 - \bar{f}_2)^2 \sum_{t=1}^T \|\hat{f}_2\|^2 x_{i,t}^4}, \\
&= O_p(T^{3/2}) + O_p\left(\frac{1}{\nu_1^2 T^{1/2}}\right) + O_p\left(\frac{\dot{\nu}_1 T^{3/4}}{\nu_1}\right).
\end{aligned}$$

Note: $(f_1^0 - \bar{f}_1)$ is I -regular with respect to θ_3 .

Similarly, we have

$$\begin{aligned}
& 2 \sum_{t=1}^T (f_2^0 - \hat{f}_2)(f_1^0 - \hat{f}_1) x_{i,t}^2 \\
&= 2 \sum_{t=1}^T (f_2^0 - \bar{f}_2 + \bar{f}_2 - \hat{f}_2)(f_1^0 - \bar{f}_1 + \bar{f}_1 - \hat{f}_1) x_{i,t}^2, \\
&= 2 \sum_{t=1}^T (f_2^0 - \bar{f}_2)(f_1^0 - \bar{f}_1) x_{i,t}^2 + 2 \sum_{t=1}^T (f_2^0 - \bar{f}_2)(\bar{f}_1 - \hat{f}_1) x_{i,t}^2 \\
&\quad + 2 \sum_{t=1}^T (\bar{f}_2 - \hat{f}_2)(f_1^0 - \bar{f}_1) x_{i,t}^2 + 2 \sum_{t=1}^T (\bar{f}_2 - \hat{f}_2)(\bar{f}_1 - \hat{f}_1) x_{i,t}^2, \\
&\leq O_p(T^{3/2}) + 2\|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\| \sum_{t=1}^T \|\hat{f}_1\| (f_2^0 - \bar{f}_2) x_{i,t}^2 \\
&\quad + 2\|\hat{\theta}_{2,n,T} - \theta_{2,0}\| \sum_{t=1}^T \|\hat{f}_2\| (f_1^0 - \bar{f}_1) x_{i,t}^2 \\
&\quad + 2\|(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' - (\theta'_{1,0}, \theta'_{2,0})'\| \|\hat{\theta}_{2,n,T} - \theta_{2,0}\| \sum_{t=1}^T \|\hat{f}_1\| \|\hat{f}_2\| x_{i,t}^2, \\
&= O_p(T^2 \nu_2^2) + O_p\left(\frac{\dot{\nu}_1 T^{3/4}}{\nu_1}\right) + O_p\left(\frac{T^{1/2}}{\nu_1}\right) + O_p\left(\frac{\dot{\nu}_1 T^{3/4}}{\nu_1}\right).
\end{aligned}$$

Hence, we have

$$M_{2,n,T}^{-1} R_{n,T,1}^* (\hat{\theta}_{1,n,T}, \hat{\theta}_{2,n,T}, \theta_3, \theta_4, \theta_0) \xrightarrow{P} \int_{-\infty}^{\infty} [\bar{r}_1 - r_1^0 + \bar{f}_2 - f_2^0]^2 ds^2 \mathbf{E} \left[\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right],$$

by the limit of

$$M_{2,n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T [(f_2^0 - \bar{f}_2)^2 + (f_1^0 - \bar{f}_1)^2 + 2(f_2^0 - \bar{f}_2)(f_1^0 - \bar{f}_1)] x_{i,t}^2,$$

where $h_1(\cdot; \theta_{1,0}, \theta_{2,0})$ cancels out in both f_1^0 and \bar{f}_1 , and only r_1^0 and \bar{r}_1 remain. Note: If θ_3 appears in h_1 , f_1 is ν_1 H -regular with respect to θ_3 but not ν_3 . \square

3.C.18 Proof of Theorem 3.7

According to Lemma 3.5, we only need to verify condition (a)-(e). For simplicity, we define $\dot{f}_{i,j} = \frac{\partial f_i}{\partial \theta_j}$, for $i = 1, 2$ and $j = 1, 2, 3, 4$, and $\dot{f}_{i,j,0}$ as there variables evaluated at θ_0 .

Condition (a): According to the definition of $g(z_{i,t}, x_{i,t}, \theta)$, $Q_{n,T}(\theta)$ and the model specification (3.1), we have

$$\dot{g}(z_{i,t}, x_{i,t}, \theta_0) = \begin{bmatrix} \dot{f}_{1,1,0} \\ \dot{f}_{1,2,0} + \dot{f}_{2,2,0} \\ \dot{f}_{1,3,0} + \dot{f}_{2,3,0} \\ \dot{f}_{2,4,0} \end{bmatrix} x_{i,t}, \quad (3.C.38)$$

$$\dot{Q}_{n,T}(\theta_0) = - \sum_{i=1}^n \sum_{t=1}^T \begin{bmatrix} \dot{f}_{1,1,0} \\ \dot{f}_{1,2,0} + \dot{f}_{2,2,0} \\ \dot{f}_{1,3,0} + \dot{f}_{2,3,0} \\ \dot{f}_{2,4,0} \end{bmatrix} x_{i,t} u_{i,t}, \quad (3.C.39)$$

and,

$$\begin{aligned} \ddot{Q}_{n,T}^0(\theta_0) &= \sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta_0) \dot{g}(z_{i,t}, x_{i,t}, \theta_0)' \\ &= \sum_{i=1}^n \sum_{t=1}^T \begin{bmatrix} \dot{f}_{1,1,0} \\ \dot{f}_{1,2,0} + \dot{f}_{2,2,0} \\ \dot{f}_{1,3,0} + \dot{f}_{2,3,0} \\ \dot{f}_{2,4,0} \end{bmatrix} \begin{bmatrix} \dot{f}_{1,1,0} \\ \dot{f}_{1,2,0} + \dot{f}_{2,2,0} \\ \dot{f}_{1,3,0} + \dot{f}_{2,3,0} \\ \dot{f}_{2,4,0} \end{bmatrix}' x_{i,t}^2 \\ &= \sum_{i=1}^n \sum_{t=1}^T \bar{A}_{i,t} x_{i,t}^2, \end{aligned} \quad (3.C.40)$$

where

$$\bar{A}_{i,t} =$$

$$\begin{bmatrix} \dot{f}_{1,1,0}\dot{f}'_{1,1,0} & \dot{f}_{1,1,0}(\dot{f}_{1,2,0} + \dot{f}_{2,2,0})' & \dot{f}_{1,1,0}(\dot{f}_{1,3,0} + \dot{f}_{2,3,0})' & \dot{f}_{1,1,0}\dot{f}'_{2,4,0} \\ (\dot{f}_{1,2,0} + \dot{f}_{2,2,0})\dot{f}'_{1,1,0} & (\dot{f}_{1,2,0} + \dot{f}_{2,2,0})(\dot{f}_{1,2,0} + \dot{f}_{2,2,0})' & (\dot{f}_{1,2,0} + \dot{f}_{2,2,0})(\dot{f}_{1,3,0} + \dot{f}_{2,3,0})' & (\dot{f}_{1,2,0} + \dot{f}_{2,2,0})\dot{f}'_{2,4,0} \\ (\dot{f}_{1,3,0} + \dot{f}_{2,3,0})\dot{f}'_{1,1,0} & (\dot{f}_{1,3,0} + \dot{f}_{2,3,0})(\dot{f}_{1,2,0} + \dot{f}_{2,2,0})' & (\dot{f}_{1,3,0} + \dot{f}_{2,3,0})(\dot{f}_{1,3,0} + \dot{f}_{2,3,0})' & (\dot{f}_{1,3,0} + \dot{f}_{2,3,0})\dot{f}'_{2,4,0} \\ \dot{f}_{2,4,0}\dot{f}'_{1,1,0} & \dot{f}_{2,4,0}(\dot{f}_{1,2,0} + \dot{f}_{2,2,0})' & \dot{f}_{2,4,0}(\dot{f}_{1,3,0} + \dot{f}_{2,3,0})' & \dot{f}_{2,4,0}\dot{f}'_{2,4,0} \end{bmatrix}.$$

Now, we define the following diagonal matrix

$$C_{n,T} = \begin{bmatrix} \sqrt{n}(T\dot{\nu}_1) & 0 & 0 & 0 \\ 0 & \sqrt{n}(T\dot{\nu}_1) & 0 & 0 \\ 0 & 0 & \sqrt{n}(T\dot{\nu}_2) & 0 \\ 0 & 0 & 0 & \sqrt{n}(T\dot{\nu}_2) \end{bmatrix},$$

where the dimension of the nonzero diagonal is comfortable with the dimensions of θ_1 , θ_2 , θ_3 and θ_4 respectively.

Then, we have

$$\begin{aligned} L_{n,T} &= -C_{n,T}^{-1}\dot{Q}_{n,T}(\theta_0) \\ &= \begin{bmatrix} \frac{1}{\sqrt{n}(T\dot{\nu}_1)} \sum_{i=1}^n \sum_{t=1}^T \dot{f}_{1,1,0}x_{i,t}u_{i,t} \\ \frac{1}{\sqrt{n}(T\dot{\nu}_1)} \sum_{i=1}^n \sum_{t=1}^T [\dot{f}_{1,2,0} + \dot{f}_{2,2,0}]x_{i,t}u_{i,t} \\ \frac{1}{\sqrt{n}(T\dot{\nu}_2)} \sum_{i=1}^n \sum_{t=1}^T [\dot{f}_{1,3,0} + \dot{f}_{2,3,0}]x_{i,t}u_{i,t} \\ \frac{1}{\sqrt{n}(T\dot{\nu}_2)} \sum_{i=1}^n \sum_{t=1}^T \dot{f}_{2,4,0}x_{i,t}u_{i,t} \end{bmatrix}. \end{aligned}$$

According to Theorem 3.3 from Park and Phillips (2001) and $\frac{\dot{\nu}_2}{\dot{\nu}_1} \xrightarrow{P} 0$, we have

$$\begin{bmatrix} \frac{1}{T\dot{\nu}_1} \sum_{t=1}^T \dot{f}_{1,1,0}x_{i,t}u_{i,t} \\ \frac{1}{T\dot{\nu}_1} \sum_{t=1}^T [\dot{f}_{1,2,0} + \dot{f}_{2,2,0}]x_{i,t}u_{i,t} \\ \frac{1}{T\dot{\nu}_2} \sum_{t=1}^T [\dot{f}_{1,3,0} + \dot{f}_{2,3,0}]x_{i,t}u_{i,t} \\ \frac{1}{T\dot{\nu}_2} \sum_{t=1}^T \dot{f}_{2,4,0}x_{i,t}u_{i,t} \end{bmatrix} \xrightarrow{d} \int_0^1 \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix} B_{x,i}dB_{u,i},$$

as $T \rightarrow \infty$. By applying CLT, for $n \rightarrow \infty$,

$$L_{n,T} \xrightarrow{d} \mathbf{N}(0, \Sigma_{HH}),$$

with

$$\Sigma_{HH} = \mathbf{E} \left\{ \int_0^1 \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix} \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix}' B_{x,i}^2 dr \right\}.$$

The above expectation results from the IID property of

$$\int_0^1 \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix} \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix}' B_{x,i}^2 dr,$$

across i , which results from Assumptions 3.1' and 3.2 and where i represents the cross-section unit. Next, we establish the convergence of $J_{n,T}$. According to Theorem 3.3 from Park and Phillips (2001) and LLNs, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{P} \mathbf{E} \left\{ \int_0^1 \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix} \begin{bmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \\ \dot{h}_{r_1,3,0} + \dot{h}_{2,3,0} \\ \dot{h}_{2,4,0} \end{bmatrix}' B_{x,i}^2 dr \right\} = \Sigma_{HH}.$$

Hence, the results of Theorem 3.7 follow from the convergence of

$$C_{n,T}(\hat{\theta}_{n,T} - \theta_0) = C_{n,T} J_{n,T}^{-1} L_{n,T}.$$

Condition (b): We define $\ddot{f}_{i,jk} = \frac{\partial^2 f_i}{\partial \theta_j \partial \theta_k}$, for $i = 1, 2$ and $j, k = 1, 2, 3, 4$, and evaluate $\ddot{f}_{i,jk,0}$ at θ_0 . According to Theorem 3.3 from Park and Phillips (2001),

$$\begin{aligned} & (\ddot{C}_{2,T} T)^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t} \\ &= (\ddot{C}_{2,T} T)^{-1} \sum_{t=1}^T \begin{bmatrix} \ddot{f}_{1,11,0} & \ddot{f}_{1,12,0} & \ddot{f}_{1,13,0} & 0 \\ \ddot{f}_{1,21,0} & \ddot{f}_{1,22,0} + \ddot{f}_{2,22,0} & \ddot{f}_{1,23,0} + \ddot{f}_{2,23,0} & \ddot{f}_{2,24,0} \\ \ddot{f}_{1,31,0} & \ddot{f}_{1,32,0} + \ddot{f}_{2,32,0} & \ddot{f}_{1,33,0} + \ddot{f}_{2,33,0} & \ddot{f}_{2,34,0} \\ 0 & \ddot{f}_{2,42,0} & \ddot{f}_{2,43,0} & \ddot{f}_{2,44,0} \end{bmatrix} x_{i,t} u_{i,t}, \\ &= O_p(1), \end{aligned}$$

where

$$\ddot{C}_{2,T} = \begin{bmatrix} \ddot{\nu}_1 & 0 & 0 & 0 \\ 0 & \ddot{\nu}_1 & 0 & 0 \\ 0 & 0 & \ddot{\nu}_2 & 0 \\ 0 & 0 & 0 & \ddot{\nu}_2 \end{bmatrix}.$$

Hence,

$$\begin{aligned}
& (C_{1,T}T)^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t} \\
&= (C_{1,T}^{-1} \ddot{C}_{2,T}) (\ddot{C}_{2,T} T)^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t}, \\
&\leq \|C_{1,T}^{-1} \ddot{C}_{2,T}\| (\ddot{C}_{2,T} T)^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t}, \\
&= O_p(1),
\end{aligned} \tag{3.C.41}$$

where $C_{1,T}$ is defined as C_T in Theorem 3.7. The above inequality follows from the fact that $C_{1,T}$ and $\ddot{C}_{2,T}$ are strictly positive. Furthermore, $\|C_{1,T}^{-1} \ddot{C}_{2,T}\| < \infty$ is due to condition (a) of Assumption H and condition (b) of Theorem 3.7. Hence,

$$\begin{aligned}
R_{n,T} &= C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta_0) u_{i,t} C_{n,T}^{-1}, \\
&= o_p(1),
\end{aligned}$$

where $C_{n,T} = \sqrt{n}TC_{1,T}$, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$.

Condition(c): For $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{p} J,$$

follows from the proof of Condition (a). Hence, $J > 0$ results from condition (c) of Theorem 3.7 and the Reverse Hölder inequality of order 2.

Condition (d): Follows from the same argument in the proof of Theorem 3.3.

Condition (e): The arguments are similar to those for the proof of Theorem 3.4. Hence, we need the following inequalities

$$\begin{aligned}
& (C_{1,T} \otimes C_{1,T})^{-1} \\
&= (C_{1,T} \otimes C_{1,T})^{-1} C_\nu \ddot{C}_{2,T} \ddot{C}_{2,T}^{-1} C_\nu^{-1}, \\
&\leq \|(C_{1,T} \otimes C_{1,T})^{-1} C_\nu \ddot{C}_{2,T}\| \ddot{C}_{2,T}^{-1} C_\nu^{-1},
\end{aligned}$$

where

$$C_\nu = \begin{bmatrix} \nu_1 & 0 & 0 & 0 \\ 0 & \nu_1 & 0 & 0 \\ 0 & 0 & \nu_2 & 0 \\ 0 & 0 & 0 & \nu_2 \end{bmatrix},$$

Condition (a) of Assumption H and condition (b) of Theorem 3.3 imply the boundedness of $\|(C_{1,T} \otimes C_{1,T})^{-1} C_\nu \ddot{C}_{2,T}\|$.

□

3.C.19 Proof of Theorem 3.8

According to Lemma 3.5, we only need to verify conditions (a)-(e).

Condition (a): $\dot{Q}_{n,T}(\theta_0)$ and $\ddot{Q}_{n,T}^0(\theta_0)$ are defined similarly to those in the proof of Theorem 3.7. Furthermore, we define the following diagonal matrix

$$C_{n,T} = \begin{bmatrix} \sqrt{n}(T\dot{\nu}_1) & 0 & 0 & 0 \\ 0 & \sqrt{n}(T\dot{\nu}_1) & 0 & 0 \\ 0 & 0 & \sqrt{n}T^{3/4} & 0 \\ 0 & 0 & 0 & \sqrt{n}T^{3/4} \end{bmatrix},$$

where the dimension of the nonzero diagonal is comfortable with the dimensions of $\theta_1, \theta_2, \theta_3$ and θ_4 respectively.

Then, we have

$$\begin{aligned} L_{n,T} &= -C_{n,T}^{-1} \dot{Q}_{n,T}(\theta_0), \\ &= \begin{bmatrix} \frac{1}{\sqrt{n}(T\dot{\nu}_1)} \sum_{i=1}^n \sum_{t=1}^T \dot{f}_{1,1,0} x_{i,t} u_{i,t} \\ \frac{1}{\sqrt{n}(T\dot{\nu}_1)} \sum_{i=1}^n \sum_{t=1}^T [\dot{f}_{1,2,0} + \dot{f}_{2,2,0}] x_{i,t} u_{i,t} \\ \frac{1}{\sqrt{n}T^{(3/4)}} \sum_{i=1}^n \sum_{t=1}^T [\dot{r}_{1,3,0} + \dot{f}_{2,3,0}] x_{i,t} u_{i,t} \\ \frac{1}{\sqrt{n}T^{(3/4)}} \sum_{i=1}^n \sum_{t=1}^T \dot{f}_{2,4,0} x_{i,t} u_{i,t} \end{bmatrix}. \end{aligned}$$

According to Theorem 7 from Chang et al. (2001) and Lemma 3.1, we have

$$\begin{aligned} & \begin{bmatrix} \frac{1}{T\nu_1} \sum_{t=1}^T \dot{f}_{1,1,0} x_{i,t} u_{i,t} \\ \frac{1}{T\nu_1} \sum_{t=1}^T [\dot{f}_{1,2,0} + \dot{f}_{2,2,0}] x_{i,t} u_{i,t} \\ \frac{1}{T^{3/4}} \sum_{t=1}^T [\dot{r}_{1,3,0} + \dot{f}_{2,3,0}] x_{i,t} u_{i,t} \\ \frac{1}{T^{3/4}} \sum_{t=1}^T \dot{f}_{2,4,0} x_{i,t} u_{i,t} \end{bmatrix} \\ & \xrightarrow{d} \begin{bmatrix} \int_0^1 \begin{pmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \end{pmatrix} B_{x,i} dB_{u,i} \\ \left\{ \int_{-\infty}^{\infty} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix}' ds \int_0^1 B_{x,i}^2 dL_{z,i}(r,0) \right\}^{1/2} W_i \end{bmatrix}, \end{aligned}$$

as $T \rightarrow \infty$, where W_i is defined in Lemma 3.1. By applying CLT, for $n \rightarrow \infty$,

$$L_{n,T} \xrightarrow{d} \mathbf{N}(0, \Sigma_{HI}),$$

with

$$\Sigma_{HI} =$$

$$\mathbf{E} \begin{bmatrix} \int_0^1 \begin{pmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \end{pmatrix} \begin{pmatrix} \dot{h}_{1,1,0} \\ \dot{h}_{1,2,0} \end{pmatrix}' B_{x,i}^2 dr & 0 \\ 0 & \int_{-\infty}^{\infty} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix}' ds \int_0^1 B_{x,i}^2 dL_{z,i}(r,0) \end{bmatrix}.$$

Next, we establish the convergence of $J_{n,T}$. According to Theorem 7 from Chang et al. (2001) and LLNs, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{P} \Sigma_{HI}.$$

Hence, the results of Theorem 3.8 follow from the convergence of

$$C_{n,T}(\hat{\theta}_{n,T} - \theta_0) = C_{n,T} J_{n,T}^{-1} L_{n,T}.$$

Condition (b): We define $\ddot{f}_{i,jk} = \frac{\partial^2 f_i}{\partial \theta_j \partial \theta_k}$, for $i = 1, 2$ and $j, k = 1, 2, 3, 4$, and $\ddot{f}_{i,jk,0}$ as these variables are

evaluated at θ_0 . According to Theorem 3.3 from Park and Phillips (2001),

$$\begin{aligned}
 & \ddot{C}_{2,T}^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t} \\
 &= \ddot{C}_{2,T}^{-1} \sum_{t=1}^T \begin{bmatrix} \ddot{f}_{1,11,0} & \ddot{f}_{1,12,0} & \ddot{f}_{1,13,0} & 0 \\ \ddot{f}_{1,21,0} & \ddot{f}_{1,22,0} + \ddot{f}_{2,22,0} & \ddot{f}_{1,23,0} + \ddot{f}_{2,23,0} & \ddot{f}_{2,24,0} \\ \ddot{f}_{1,31,0} & \ddot{f}_{1,32,0} + \ddot{f}_{2,32,0} & \ddot{f}_{1,33,0} + \ddot{f}_{2,33,0} & \ddot{f}_{2,34,0} \\ 0 & \ddot{f}_{2,42,0} & \ddot{f}_{2,43,0} & \ddot{f}_{2,44,0} \end{bmatrix} x_{i,t} u_{i,t}, \\
 &= O_p(1),
 \end{aligned}$$

where

$$\ddot{C}_{2,T} = \begin{bmatrix} T\ddot{\nu}_1 & 0 & 0 & 0 \\ 0 & T\ddot{\nu}_1 & 0 & 0 \\ 0 & 0 & T^{3/4} & 0 \\ 0 & 0 & 0 & T^{3/4} \end{bmatrix}.$$

Hence,

$$\begin{aligned}
 & C_{1,T}^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t} \\
 &= (C_{1,T}^{-1} \ddot{C}_{2,T}) \ddot{C}_{2,T}^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t}, \\
 &\leq \|C_{1,T}^{-1} \ddot{C}_{2,T}\| (\ddot{C}_{2,T} T)^{-1} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} u_{i,t}, \\
 &= O_p(1),
 \end{aligned} \tag{3.C.42}$$

where

$$C_{1,T} = \begin{bmatrix} T\dot{\nu}_1 & 0 & 0 & 0 \\ 0 & T\dot{\nu}_1 & 0 & 0 \\ 0 & 0 & T^{3/4} & 0 \\ 0 & 0 & 0 & T^{3/4} \end{bmatrix}.$$

The above inequality follows from the fact that $C_{1,T}$ and $\ddot{C}_{2,T}$ are strictly positive. Furthermore, $\|C_{1,T}^{-1} \ddot{C}_{2,T}\| < \infty$ is due to condition (a) of Assumption H. Hence,

$$\begin{aligned}
 R_{n,T} &= C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta_0) u_{i,t} C_{n,T}^{-1}, \\
 &= o_p(1),
 \end{aligned}$$

where $C_{n,T} = \sqrt{n}C_{1,T}$, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$.

Condition(c): For $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{p} J,$$

follows from the proof of Condition (a). Hence, $J > 0$ results from condition (b) of Theorem 3.8 and the Reverse Hölder inequality of order 2.

Condition (d): Follows from the same argument in the proof of Theorem 3.3.

Condition (e): According to Theorem 7 from Chang et al. (2001), Lemma 3.1 and part (e) in the proof of Theorem 3.3 and Theorem 3.7, we have

$$\begin{aligned} C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta) \dot{g}(z_{i,t}, x_{i,t}, \theta)' C_{n,T}^{-1} &\xrightarrow{d} J(\theta), \\ C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta) [g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0)] C_{n,T}^{-1} &= o_p(1), \\ C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta) u_{i,t} C_{n,T}^{-1} &= o_p(1), \end{aligned}$$

uniformly in $\theta \in \Theta$, where $J(\theta)$ is obtained by replacing θ_0 of J by θ . Hence, if $\theta_{n,T}$ converges to θ_0 following from $\hat{\theta}_{n,T} \xrightarrow{p} \theta_0$, we have

$$C_{n,T}^{-1} (\ddot{Q}_{n,T}(\theta_{n,T})) C_{n,T}^{-1} \xrightarrow{p} J, \quad (3.C.43)$$

for $T \rightarrow \infty$ followed by $n \rightarrow \infty$. Besides, using the proofs of Condition (a) and Condition (b), we get

$$C_{n,T}^{-1} (\ddot{Q}_{n,T}(\theta_0)) C_{n,T}^{-1} \xrightarrow{p} J.$$

Hence, Condition (e) is established in view of above convergence and (3.C.43). \square

3.C.20 Proof of Corollary 3.3

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 3.5, the results in Corollary 3.3 is a natural byproduct of the proof of Theorem 3.7 in view of Condition (e) in Lemma 3.5. \square

3.C.21 Proof of Corollary 3.4

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 3.6, the results in Corollary 3.4 is a natural byproduct of the proof of Theorem 3.8 in view of Condition (e) in Lemma 3.5. \square

3.C.22 Proof of Corollary 3.5

Conditions (b) and (c) in Lemma 3.3 follow from the same arguments as those in the proof of Theorem 3.1. We only need to define $M_{n,T}$, $R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0)$ and $R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0)$, and verify the following uniform convergence

$$M_{n,T}^{-1} R_{n,T,1}^*(\theta, \theta_0) \xrightarrow{p} R^*(\theta, \theta_0), \quad (3.C.44)$$

$$M_{n,T}^{-1} R_{n,T,2}^*(\theta, \theta_0) \xrightarrow{p} 0, \quad (3.C.45)$$

where $\theta = (\alpha, \beta, \gamma, c)'$. According to (3.23) in Section 3.5 and (3.10) in Section 3.3, we have

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0 f_3^0 + f_2^0 f_4^0) - (f_1 + f_2 f_3 + f_2 f_4)]^2,$$

and,

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 + f_2^0 f_3^0 + f_2^0 f_4^0) - (f_1 + f_2 f_3 + f_2 f_4)] u_{i,t}.$$

We take $M_{n,T} = nT^2$ and establish (3.C.45) first. According to Lemma 5(d) from Chang et al. (2001), we have

$$\sum_{t=1}^T (f_1^0 - f_1) u_{i,t} = O_p(T).$$

Hence,

$$M_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T (f_1^0 - f_1) u_{i,t} = o_p(1),$$

as $T \rightarrow \infty$ followed by $n \rightarrow \infty$. According to Lemma 3.1, we have

$$\sum_{t=1}^T (f_2^0 f_3^0 - f_2 f_3) u_{i,t} = O_p(T^{3/4}).$$

Hence,

$$M_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T (f_2^0 f_3^0 - f_2 f_3) u_{i,t} = o_p(1).$$

$$\begin{aligned}
& \mathbf{E} \left[\frac{1}{T} \sum_{t=1}^T (f_2^0 f_4^0 - f_2 f_4) u_{i,t} \right]^2 \\
&= \mathbf{E} \sum_{t=1}^T \frac{1}{T} [(\beta_0 f_4^0 - \beta f_4)^2 \left(\frac{x_{i,t}}{\sqrt{T}} \right)^2 u_{i,t}^2], \\
&\leq \mathbf{E} \sum_{t=1}^T \frac{1}{T} [B_\beta^2 \left(\frac{x_{i,t}}{\sqrt{T}} \right)^2 u_{i,t}^2], \\
&= B_\beta^2 \mathbf{E}[u_{i,t}^2] \mathbf{E} \sum_{t=1}^T \frac{1}{T} \left[\left(\frac{x_{i,t}}{\sqrt{T}} \right)^2 \right], \\
&= O_p(1).
\end{aligned} \tag{3.C.46}$$

The first and second equalities follow from the assumptions of martingale difference $u_{i,t}$ and weak exogeneity. The inequality holds because $-1 \leq f_4 \leq 0$. Accordingly, $|\beta_0 f_4^0 - \beta f_4|$ is bounded by a positive number B_β given the compactness of the parameter space. Hence,

$$M_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T (f_2^0 f_4^0 - f_2 f_4) u_{i,t} = o_p(1).$$

Therefore, (3.C.45) holds.

Now, we try to establish (3.C.44).

$$\begin{aligned}
R_{n,T,1}^*(\theta, \theta_0) &= \sum_{i=1}^n \sum_{t=1}^T [(f_1^0 - f_1)^2 + (f_2^0 f_3^0 - f_2 f_3)^2 + (f_2^0 f_4^0 - f_2 f_4)^2 \\
&\quad + (f_1^0 - f_1)(f_2^0 f_3^0 - f_2 f_3) + 2(f_1^0 - f_1)(f_2^0 f_4^0 - f_2 f_4) + 2(f_2^0 f_3^0 - f_2 f_3)(f_2^0 f_4^0 - f_2 f_4)].
\end{aligned}$$

According to Lemma 3.1, $\sum_{t=1}^T (f_2^0 f_3^0 - f_2 f_3)^2$ and $\sum_{t=1}^T [2(f_1^0 - f_1)(f_2^0 f_3^0 - f_2 f_3)]$ are both $O_p(T^{3/2})$. Furthermore, $\sum_{t=1}^T [2(f_2^0 f_3^0 - f_2 f_3)(f_2^0 f_4^0 - f_2 f_4)]$ is also $O_p(T^{3/2})$. The convergence follows from the same arguments as those for (3.C.46). Hence, they converge to zero once being scaled by $M_{n,T} = nT^2$. Furthermore, we have

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T (f_1^0 - f_1)^2 \xrightarrow{d} [\phi_0 - \phi]^2 \int_0^1 B_{x,i}^2(r) dr, \\
& \frac{1}{T^2} \sum_{t=1}^T (f_2^0 f_4^0 - f_2 f_4)^2 \xrightarrow{d} \int_0^1 (\beta_0 - \beta)^2 1\{B_{z,i} > 0\} B_{x,i}^2(r) dr, \\
& \frac{1}{T^2} \sum_{t=1}^T [2(f_1^0 - f_1)(f_2^0 f_4^0 - f_2 f_4)] \xrightarrow{d} 2[\phi_0 - \phi] \int_0^1 (\beta_0 - \beta) 1\{B_{z,i} > 0\} B_{x,i}^2(r) dr.
\end{aligned}$$

The above convergence results follow from the continuous mapping theorem and Lemma 5 from Chang et al. (2001). Moreover, we use Lemma A2 with $f(x_{i,t}) = x_{i,t}^2$ being $S(5/2)$ to have

$$\frac{1}{T^2} \sum_{t=1}^T [\beta_0 f_4^0 - \beta_0 1\{z_{i,t} > 0\} (1 - \frac{2}{\exp(\gamma_0(z_{i,t} - c_0)) + 1})] = o_p(1),$$

and

$$\frac{1}{T^2} \sum_{t=1}^T [\beta f_4 - \beta 1\{z_{i,t} > 0\} (1 - \frac{2}{\exp(\gamma(z_{i,t} - c)) + 1})] = o_p(1).$$

The remainder term of f_4 can be proved to be $o_p(T^2)$ by the same arguments as those for (3.C.46). Combining the above three convergence results and applying LLNs as $n \rightarrow \infty$, we have

$$R^*(\theta, \theta_0) = \mathbf{E}[\int_0^1 [(\phi_0 - \phi) + (\beta_0 - \beta) 1\{B_{z,i} > 0\}]^2 B_{x,i}^2(r) dr],$$

The uniform convergence follows from Lemma 3.C.2 and the same arguments as those in the proof of Theorem 3.1 and Theorem 3.2. □

3.C.23 Proof of Corollary 3.6

Following the proof of Theorem 3.6. □

3.C.24 Proof of Corollary 3.7

Following the proof of Theorem 3.8. □

3.C.25 Proof of Corollary 3.8

Following the proof of Theorem 3.8. □

Nonlinear Least Squares Dummy Variable Estimation for Independent Co-summable Panels with Fixed Effect

4.1 Introduction

In the econometric analysis of linear panel data models, components such as fixed effects or random effects are introduced into the model to capture the possibly unobservable heterogeneity across individuals within the panel. Correspondingly, different estimators are proposed, i.e., the Least Squares Dummy Variable (LSDV) estimator for models with fixed effects. The asymptotic properties of such estimators are well studied and understood in the econometric literature among which Baltagi (2008) contains the most recent overview of the theory. Although nonlinearity introduces extra flexibility in modeling panel data, the fixed effects or random effects are still expected to be capable of capturing the unobserved or unexplained heterogeneity in many econometric applications. A recent example of nonlinear panel models with fixed effects can be found in González et al. (2005) and Fok, Van Dijk and Franses (2005), where a forward combination of the linear fixed effect estimator and the NLS estimator is proposed without detailed theoretical analysis.

In this chapter, we extend the nonlinear panel co-summable model in Chapter 3 by introducing an additive fixed effect term for each i . Moreover, we formalize the estimator used by González et al. (2005) as the Pooled Nonlinear Least Squares Dummy Variable (NLSDV) estimator, and derive its asymptotic properties for panel simple strong co-summability and panel multiple strong co-summability respectively. The results show that the Pooled NLSDV estimator has the same consistency rate as the Pooled NLS estimator for the same model without fixed effects as $T \rightarrow \infty$ being followed by $n \rightarrow \infty$. The Pooled NLSDV estimator requires a set of different identification conditions, which might be also stronger than those required by the Pooled NLS estimator. Similar to the linear case, an extra term due to the demeaning appears in the asymptotic distribution if the nonlinear function is H -regular. Different from the linear case, the asymptotic distribution is not different from that of the Pooled NLS estimator if the nonlinear function is I -regular.

The rest of the chapter is organized as follows. Section 4.2 lays out the basic model and assumptions, and introduces the Pooled NLSDV estimator. In Section 4.3, the asymptotic results are derived for panel simple strong co-summable regression models including I -regular and H -regular functions. Section 4.4 contains the consistency and asymptotic distribution results for panel multiple strong co-summable regression models. Section 4.5 takes the Logistic Smooth Transition Co-summable Regression (LSTCR) model as an example, and proposes several estimators for the covariance matrix. In Section 4.6, the simulation study is conducted based on the LSTCR model. Moreover, the performance of different estimators for the covariance matrix is compared for inference purposes. Some concluding remarks are made in Section 4.7.

A word about the notations. For a vector $x = (x_i)$ or a matrix $A = (a_{ij})$, the modulus $|\cdot|$ is taken element by element. Therefore, $|x| = (|x_i|)$ and $|A| = (|a_{ij}|)$. The standard Euclidean norm of a vector is denoted by $\|\cdot\|$, i.e. $\|x\|^2 = \sum_i (x_i)^2$. For a matrix, $\|A\|$ signifies the operator norm defined by $\|A\| = \sup_x \|Ax\|/\|x\|$. The maximum of the moduli is denoted by $\overline{\|\cdot\|}$, i.e., $\overline{\|x\|} = \max_i |x_i|$ and $\overline{\|A\|} = \max_{i,j} |a_{ij}|$. Moreover, the k by k dimension identity matrix is denoted by \mathbf{I}_k . For a function, which can be vector- or matrix-valued, $\overline{\|\cdot\|}_K$ signifies the supremum over a subset K of its domain, so that $\overline{\|f\|}_K = \sup_{x \in K} \|f(x)\|$. The subset K , over which the supremum is taken, will not be specified if it is clear from the context. Furthermore, the indicator function is written as $1\{\cdot\}$. Letters such as p, q, m, l and δ generally denote positive integers and the definitions should be clear under the context. $=_d$ denotes equality in distribution. In addition, standard terminologies and notations in probability and measure theory are used throughout the rest of the chapter such as almost sure convergence ($\xrightarrow{a.s.}$), convergence in probability (\xrightarrow{P}), convergence in distribution (\xrightarrow{d}) and weak convergence (\Rightarrow). $W(r)$ denotes a vector of standard Brownian motions and $B(r)$ denotes a vector of general Brownian motions. Their dimensions are clear from the context. For a function $f(x, \theta)$, \dot{f} , \ddot{f} and \dddot{f} denote, respectively, the first, second and third derivatives of f with respect to θ . They are possibly vectors and arranged by the lexicographic ordering of their indices based on the vector θ .

4.2 The Model and Assumptions

We consider the following cross-sectionally independent panel nonlinear regression model for a scalar endogenous variable $y_{i,t}$ given by

$$y_{i,t} = \delta_i + f(z_{i,t}, \theta)x_{i,t} + u_{i,t}, \quad (4.2.1)$$

where θ is a finite dimension vector of parameters, $z_{i,t}$ and $x_{i,t}$ are scalar regressors assumed to be generated by

$$\begin{bmatrix} x_{i,t} \\ z_{i,t} \end{bmatrix} = \begin{bmatrix} x_{i,t-1} \\ z_{i,t-1} \end{bmatrix} + \begin{bmatrix} v_{x,i,t} \\ v_{z,i,t} \end{bmatrix}, \quad (4.2.2)$$

with common initialization at $t = 0$ satisfying

$$(z_{i,0}, x_{i,0})' \text{ being IID across } i \text{ with } \mathbf{E}\|(z_{i,0}, x_{i,0})'\|^4 < \infty.$$

The innovation term $V_{i,t} = (v_{x,i,t}, v_{z,i,t})'$ is assumed to be generated by the linear process with random coefficients

$$V_{i,t} = \varphi_i(L)\eta_{i,t} = \sum_{s=0}^{\infty} \varphi_{i,s} \begin{bmatrix} \varepsilon_{x,i,t-s} \\ \varepsilon_{z,i,t-s} \end{bmatrix}, \quad (4.2.3)$$

where $\varphi_{i,s}$ are 2 by 2 matrices. The function $f(\cdot, \theta)$ is assumed to be all known but not θ and identical across sections. Besides, the parameters to be estimated, θ , are also identical across sections. Following Park and Phillips (2001), we define $\xi_{i,t} = (u_{i,t}, \eta'_{i,t+1})'$ and the filtration $\mathcal{F}_{i,t} = \sigma(\{\xi_{i,s}\}_{s=-\infty}^{t-1})$, i.e., the σ -field generated by $\{\xi_{i,s}\}_{s \leq t-1}$. The following assumption is made for the innovation process $\{\xi_{i,t}\}_{t=-\infty}^{\infty}$.

Assumption 4.1. (DGP)

- (a) $(\xi_{i,t}, \mathcal{F}_{i,t})$ is a stationary and ergodic martingale difference sequence, which is independent across i ,
- (b) the vectors $\eta_{i,t}$ are IID for all i and t with $\mathbf{E}(\eta_{i,t}) = 0$ and $\mathbf{E}(\eta_{i,t}\eta'_{i,t}) = \mathbf{I}_2$,
- (c) $\mathbf{E}(\xi_{i,t}\xi'_{i,t} | \mathcal{F}_{i,t-1}) = \Sigma_i$, where Σ_i is positive definite and nonsingular, which is partitioned as

$$\Sigma_i = \begin{bmatrix} \sigma_{uu,i}^2 & \sigma_{u\varepsilon_x,i} & \sigma_{u\varepsilon_z,i} \\ \sigma_{\varepsilon_x u,i} & 1 & 0 \\ \sigma_{\varepsilon_z u,i} & 0 & 1 \end{bmatrix}, \quad (4.2.4)$$

- (d) $\mathbf{E}\|\eta_{i,t}\|^p < \infty$ for some $p > 8$,
- (e) $\sup_{t \geq 1} \mathbf{E}(\|\xi_{i,t}\|^q | \mathcal{F}_{i,t-1}) < \infty$ for some $q > 4$,

- (f) The distribution of $\eta_{i,t}$ is absolutely continuous with respect to the Lebesgue measure and has characteristic function $\Phi(\lambda)$ satisfying

$$\lim_{\|\lambda\| \rightarrow \infty} \Phi(\lambda) = o(\|\lambda\|^\delta),$$

for some $\delta > 0$.

Conditions (a) and (c)-(f) are required by exploring the time dimension properties as $T \rightarrow \infty$, which is based on the nonlinear time series regression asymptotic results developed in Chang et al. (2001). Condition (b) and the restriction on Σ_i is conventional as we assume the data generating processes of $x_{i,t}$ and $z_{i,t}$ follow a linear process, e.g., Phillips and Moon (1999b).

Assumption 4.1'. Items (a) to (f) from Assumption 4.1 hold with

$$\Sigma_i = \begin{bmatrix} \sigma_{uu,i}^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.2.5)$$

This assumption will be needed for the case where $f(z_{i,t}, \theta)$ is asymptotically homogeneous functions of $z_{i,t}$. Moreover, a set of assumptions regarding the random coefficients $\varphi_{i,s}$ are required. Let $\varphi_{a,i,s}$ be the a th element of $\text{vec}(\varphi_{i,s})$ and the k th moment $\mathbf{E}(\varphi_{a,i,s}^k) = \sigma_{k,a,s}$.

Assumption 4.2. (Random Coefficients)

- (a) $\varphi_{i,s}$ is a random matrix across i and over s , and IID across i for all s ,
- (b) $\varphi_{i,s}$ and $\eta_{j,t}$ are independent for all i, j, t , and s ,
- (c) $\mathbf{E}\|\varphi_{i,s}\|^4 < \infty$ for all s .
- (d) $\sum_{s=0}^{\infty} s^2 \sigma_{2,a,s} < \infty$,
- (e) $\sum_{s=0}^{\infty} s^4 (\sigma_{4,a,s})^{1/4} < \infty$,
- (f) $\varphi_i(1)\varphi_i(1)'$ is almost surely positive definite,

Conditions (a)-(e) are used by Phillips and Moon (1999b) for studying linear nonstationary panel models. Moreover, according to Lemma 1 of Phillips and Moon (1999a), Conditions (d) and (e) are

sufficient for $\varphi_i(1) = \sum_{s=0}^{\infty} \varphi_{i,s} < \infty$ a.s.. Condition (f) makes sure the covariance matrix of $x_{i,t}$ and $z_{i,t}$ are well defined. Moreover, Condition (f) ensures that $x_{i,t}$ and $z_{i,t}$ are random walk processes. According to Lemma 2 from Phillips and Moon (1999b), under Assumptions 4.1(4.1') and 4.2, the process $V_{i,t}$ defined in (4.2.3) admits the following Beveridge-Nelson (BN) decomposition,

$$V_{i,t} = \varphi_i(1)\eta_{i,t} + \tilde{V}_{i,t-1} - \tilde{V}_{i,t} \text{ a.s.,}$$

where $\tilde{V}_{i,t}$ are well defined square integrable random vectors. Hence, the partial sums of $V_{i,t}$ can be written as

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} V_{i,t} = \varphi_i(1) \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \eta_{i,t} + \frac{1}{\sqrt{T}} \tilde{V}_{i,0} - \frac{1}{\sqrt{T}} \tilde{V}_{i,[Tr]} \text{ a.s.,} \quad (4.2.6)$$

where $[Tr]$ denotes the integer part of Tr . Furthermore, we define

$$B_{u,i,T}^0(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_{i,t} \quad \text{and} \quad B_{v,i,T}^0(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} V_{i,t}.$$

According to Phillips and Solo (1992) and the panel functional central limit theorem (CLT) asymptotics from Phillips and Moon (1999b), the following holds for the vector Brownian motion process $(B_{u,i}, B'_{v,i})'$,

$$(B_{u,i,T}^0(r), B_{v,i,T}^{0r}(r))' \xrightarrow{d} (B_{u,i}, B'_{v,i})',$$

with both $(B_{u,i,T}^0(r), B_{v,i,T}^{0r}(r))'$ and $(B_{u,i}, B'_{v,i})'$ being defined in $D[0,1]^3$, where $D[0,1]$ is the space of cadlag functions on $[0,1]$. As Park and Phillips (2001) pointed out, the above convergence in distribution can be interpreted as weak convergence in $D[0,1]^3$ with the supremum norm topologized with the uniform topology. Then, according to the Skorodhod representation theorem, there is a common probability space¹ supporting $(B_{u,i,T}, B'_{v,i,T})'$ and $(B_{u,i}, B'_{v,i})'$ such that,

$$(B_{u,i,T}^0, B_{v,i,T}^{0r})' =^d (B_{u,i,T}, B'_{v,i,T})' \quad \text{and} \quad (B_{u,i,T}, B'_{v,i,T})' \xrightarrow{a.s.} (B_{u,i}, B'_{v,i})',$$

in $D[0,1]^3$ with the uniform topology and $B_{v,i} = (B_{x,i}, B'_{z,i})'$. The following proofs of weak consistency are all based on $(B_{u,i,T}, B'_{v,i,T})'$. Furthermore, let Θ denote the parameter space of θ such that θ_0 is the true value of the parameter. The following assumption is conventional in the nonlinear econometric literature to ensure the uniqueness and identifiability of the parameters (See, for example, Park and Phillips (2001)).

¹The original reasoning is based on assuming deterministic coefficients of the underlying moving average processes. Here, the random coefficients $\varphi_{i,t}$ are introduced with a probability space independent of above mentioned common space. Hence, the common space can be enlarged by the product measure given the independence across i .

Assumption 4.3. (Parameter Space)

- (a) Θ is compact and convex,
- (b) θ_0 is an interior point of Θ .

If the function f is H -regular, \dot{h} , \ddot{h} and \ddot{h} are, respectively, denoted as the limiting homogeneous functions of \dot{f} , \ddot{f} and \ddot{f} . Moreover, $\dot{\nu}$, $\ddot{\nu}$ and $\ddot{\nu}$ denote the corresponding asymptotic orders. Moreover, $\dot{\nu}$, $\ddot{\nu}$ and $\ddot{\nu}$ are only functions of \sqrt{T} because the random walk process is usually scaled by \sqrt{T} for deriving its asymptotic properties. With θ possibly being a vector, the derivative functions are also possibly vector functions. Hence, it is convenient to define $\dot{\nu}$, $\ddot{\nu}$ and $\ddot{\nu}$ as diagonal matrices with the asymptotic orders being the nonzero diagonal elements. In order to derive the asymptotic consistency property of the Pooled NSL estimator, we further require the following assumptions as in Chang et al. (2001).

Assumption I. *The function $f \in \mathcal{I}$, satisfies*

- (a) \dot{f} , \ddot{f} and \ddot{f} are I -regular,
- (b) $\int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds > 0$.

Here, we take Definition 3.3 from Park and Phillips (2001). It is obvious from the original definition that there is a trade off between smoothness and moments condition p .

Assumption H. *The function $f \in \mathcal{H}_0$, satisfies*

- (a) \dot{f} , \ddot{f} and \ddot{f} are H -regular with $|(\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu}| < \infty$,
- (b) $\int_{|s| \leq \delta} \dot{h}(s, \theta_0) \dot{h}(s, \theta_0)' ds > 0$ for all $\delta > 0$,

Remark 4.1.

- (1) \mathcal{I} and \mathcal{H}_0 refer to the class of integrable functions and a subset of the class of asymptotically homogeneous functions as defined in Park and Phillips (2001), respectively. To ensure \mathcal{I} and \mathcal{H}_0 are disjoint, $\liminf_{\lambda \rightarrow \infty} \nu(\lambda) > 0$ and $h \neq 0$, where $\nu(\lambda) > 0$ is the *asymptotic order* and h is the *limiting homogeneous function* as defined in Park and Phillips (2001).
- (2) The above mentioned Assumptions 4.1(4.1')-4.3, **I** and **H** are exactly the same set of assumptions used in Chapter 3.

For the simplicity of notation, let

$$g(z_{i,t}, x_{i,t}, \theta) = f(z_{i,t}, \theta)x_{i,t}. \quad (4.2.7)$$

Hence, Model (4.2.1) can be rewritten as

$$y_{i,t} = \delta_i + g(z_{i,t}, x_{i,t}, \theta) + u_{i,t}. \quad (4.2.8)$$

In order to eliminate the unobserved term δ_i , we take the time average for each i :

$$\begin{aligned} \bar{y}_i &= \delta_i + \bar{g}_i(z_{i,t}, x_{i,t}, \theta) + \bar{u}_i, \\ \frac{1}{T} \sum_{t=1}^T y_{i,t} &= \delta_i + \frac{1}{T} \sum_{t=1}^T g(z_{i,t}, x_{i,t}, \theta) + \frac{1}{T} \sum_{t=1}^T u_{i,t}. \end{aligned} \quad (4.2.9)$$

The Pooled NLSDV estimator is based on the following auxiliary model that is mathematically equal to (4.2.8) minus (4.2.9):

$$\begin{aligned} y_{i,t} - \bar{y}_i &= g(z_{i,t}, x_{i,t}) - \bar{g}_i(z_{i,t}, x_{i,t}, \theta) + u_{i,t} - \bar{u}_i, \\ y_{i,t} - \frac{1}{T} \sum_{t=1}^T y_{i,t} &= g(z_{i,t}, x_{i,t}) - \frac{1}{T} \sum_{t=1}^T g(z_{i,t}, x_{i,t}, \theta) + u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t}. \end{aligned} \quad (4.2.10)$$

Hence, the objective function for the Pooled NLSDV estimator is defined as:

$$Q_{n,T}(\theta) = \sum_{i=1}^n \sum_{t=1}^T [(y_{i,t} - \frac{1}{T} \sum_{t=1}^T y_{i,t}) - (g(z_{i,t}, x_{i,t}, \theta) - \frac{1}{T} \sum_{t=1}^T g(z_{i,t}, x_{i,t}, \theta))]^2. \quad (4.2.11)$$

The Pooled NLSDV estimator $\hat{\theta}_{n,T}$ is defined as the minimizer of $Q_{n,T}(\theta)$ over $\theta \in \Theta$,

$$\hat{\theta}_{n,T} = \arg \min_{\theta \in \Theta} Q_{n,T}(\theta). \quad (4.2.12)$$

Remark 4.2.

- (1) Eq. (4.2.10) always holds by construction, which is independent from any statistical or probabilistic properties. The basic idea is to exploit the information along the time dimension and find a new model equation that is independent of the fixed effect δ_i and retaining the parameters of interest. The parameters of interest are estimated based on the auxiliary model and their asymptotic properties are studied accordingly, provided that the parameters are identifiable

from the auxiliary model.

4.3 Simple Strong Co-summable Regressions

In this section, we provide the asymptotic properties of the Pooled NLSDV estimator for the simple co-summability relationship, defined by Definition 3.1 in Chapter 3, and the model specification (4.2.1).

4.3.1 Consistency

Similar to that of the Pooled NLS estimator, we consider the following concentrated objective function based on Eq. (4.2.11):

$$R_{n,T}^*(\theta, \theta_0) = Q_{n,T}(\theta) - Q_{n,T}(\theta_0). \quad (4.3.1)$$

The above concentrated objective function is basically the objective function for the Pooled NLS estimator of the auxiliary model. The following consistency results are obtained based on checking the conditions specified in Lemma 3.3 with Assumptions 4.1(4.1')-4.3.

Theorem 4.1. *Let Assumptions 4.1-4.3 hold, and let $f(z_{i,t}, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ satisfy Assumption H. Moreover, if Eq. (4.2.1) satisfies Definition 3.1,*

- (a) $\nu(\sqrt{T})$ is bounded away from zero as $T \rightarrow \infty$, and
- (b) $[h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]B_{x,i}(r) \neq \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]B_{x,i}(s)ds \right)$ for all $\theta \neq \theta_0$ and some $r \in [0, 1]$ almost surely,

we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$R^*(\theta, \theta_0) = \mathbf{E} \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 B_{x,i}^2(r) dr - \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)] B_{x,i}(r) dr \right)^2 \right),$$

with $M_{n,T} = nT(\nu(\sqrt{T}))^2$. Moreover,

$$\hat{\theta}_{n,T} \xrightarrow{p} \theta_0.$$

The identification condition (b) is stronger than that of the Pooled NLS estimator for panel co-summable regression models without fixed individual effects. This assumption is also used by Saikko-

nen and Choi (2004). Moreover, this condition indicates that the correlation between $B_{x,i}$ and $B_{z,i}$ is restricted, which needs to be verified upon the chosen nonlinear transformation.

Theorem 4.2. *Let Assumptions 4.1-4.3 hold, and let $f(z_{i,t}, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ satisfy Assumption I. Moreover, if Eq. (4.2.1) satisfies Definition 3.1 and $\int_{-\infty}^{\infty} [f(s, \theta) - f(s, \theta_0)]^2 ds > 0$ for all $\theta \neq \theta_0$, we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$R^*(\theta, \theta_0) = \left(\int_{-\infty}^{\infty} [f(s, \theta) - f(s, \theta_0)]^2 ds \right) \mathbf{E} \left(\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right),$$

with $M_{n,T} = nT^{3/2}$. Moreover,

$$\hat{\theta}_{n,T} \xrightarrow{p} \theta_0.$$

For the I -regular function, the fact that the Pooled NLSDV estimator having the same asymptotic limit of the concentrated objective function as the Pooled NLS estimator is consistent with the conclusion in Chang et al. (2001), which indicates that the I -regular transformation is asymptotically orthogonal to the deterministic components in the model.

4.3.2 Asymptotic Normality

According to the theoretical arguments provided in Lemma 3.5, the asymptotic distributions of the Pooled NLSDV estimator for the simple strong co-summability defined by I -regular and H -regular functions are obtained in the following theorems.

Theorem 4.3. *Let Assumptions 4.1', 4.2 and 4.3 hold, and let $f(z_{i,t}, \theta)$ satisfy the conditions in Theorem 4.1. Moreover, if*

$$\int_0^1 \left(\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i} ds \right) \left(\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i} ds \right)' dr > 0, \text{ a.s.,}$$

we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\sqrt{nT} \dot{\nu}(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{d} \mathbf{N}(0, \Sigma_H^{-1}),$$

where

$$\Sigma_H = \mathbf{E} \left[\int_0^1 \left(\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i} ds \right) \left(\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i} ds \right)' dr \right].$$

Theorem 4.3 shows that if the nonlinear function is H -regular, the asymptotic distribution of the Pooled NLSDV estimator is different from that of the Pooled NLS estimator by an extra term in the covariance matrix. Given the linear function is a special case of H -regular functions, results in Theorem 4.3 is in line with the results in the linear case.

Theorem 4.4. *Let Assumptions 4.1-4.3 hold, and let $f(z_{i,t}, \theta)$ satisfy the conditions in Theorem 4.2, we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\sqrt{n} \sqrt[3/4]{T} (\hat{\theta}_{n,T} - \theta_0) \xrightarrow{d} \mathbf{N}(0, \Sigma^{-1}),$$

where

$$\Sigma = \int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right].$$

Theorem 4.4 confirms the conclusion drawn from Theorem 4.2, which indicates that if the nonlinear function is I -regular, the asymptotic distribution of the Pooled NLSDV is independent from the fixed effect term.

4.3.3 Estimation of the Covariance Matrix

In this section, we consider the following estimator for the covariance matrices Σ and Σ_H and prove its weak consistency,

$$\hat{\Sigma}_{n,T} = C_{n,T}^{-1} \begin{bmatrix} \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\beta}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{c}_{n,t}} \\ \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\beta}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{c}_{n,t}} \\ \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{\gamma}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} \\ \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\alpha}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\beta}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{c}_{n,t}^2} \end{bmatrix} C_{n,T}^{-1}, \quad (4.3.2)$$

where $\hat{Q}_{n,T}$ is the objective function (4.2.11) being evaluated at the estimates. Then, the following corollaries show the consistency of $\hat{\Sigma}_{n,T}$.

Corollary 4.1. *Let the assumptions of Theorem 4.3 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma_H,$$

where $C_{n,T} = \sqrt{nT}\dot{\nu}$

$$\Sigma_H = \mathbf{E} \left[\int_0^1 \left(\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i} ds \right) \left(\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i} ds \right)' dr \right].$$

Corollary 4.2. *Let the assumptions of Theorem 4.4 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma,$$

where $C_{n,T} = \sqrt{n}^{3/4} \sqrt{T}$ and

$$\Sigma = \int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, \theta) \right].$$

4.4 Multiple Strong Co-summable Regressions

In this section, we consider the Pooled NLSDV estimator for the multiple strong co-summability relationship, defined by Definition 3.2 in Chapter 3, for the model specification (4.2.1). Accordingly, we further specify Model (4.2.1) as follows:

$$y_{i,t} = \delta_i + \{f_1(z_{i,t}; \theta_1, \theta_2, \theta_3) + f_2(z_{i,t}; \theta_2, \theta_3, \theta_4)\} x_{i,t} + u_{i,t}, \quad (4.4.1)$$

with $\theta = (\theta'_1, \theta'_2, \theta'_3, \theta'_4)'$ and $f(z_{i,t}; \theta) = f_1(z_{i,t}; \theta_1, \theta_2, \theta_3) + f_2(z_{i,t}; \theta_2, \theta_3, \theta_4)$. Similar to Chapter 3, we distinguish multiple strong co-summability defined by Assumption HH in Chapter 3 from that defined by Assumption HI in Chapter 3. Recall that h generally denotes the *limiting homogeneous function* of a H -regular function. For simplicity of the notation, we define $h_{r_1}^0 = h_{r_1}(s; \theta_{1,0}, \theta_{2,0}, \theta_{3,0})$, $\bar{h}_{r_1} = h_{r_1}(s; \theta_{1,0}, \theta_{2,0}, \theta_3)$, $h_2^0 = h_2(s; \theta_{2,0}, \theta_{3,0}, \theta_{4,0})$, $\bar{h}_2 = h_2(s; \theta_{2,0}, \theta_3, \theta_4)$, $\bar{h}_{r_1}^B = h_{r_1}(B_{z,i}; \theta_{1,0}, \theta_{2,0}, \theta_3)$, $h_{r_1}^{B,0} = h_{r_1}(B_{z,i}; \theta_{1,0}, \theta_{2,0}, \theta_{3,0})$, $\bar{h}_2 = h_2(B_{z,i}; \theta_{2,0}, \theta_3, \theta_4)$ and $h_2^{B,0} = h_2(B_{z,i}; \theta_{2,0}, \theta_{3,0}, \theta_{4,0})$, where h_{r_1} is the homogeneous function of the remainder term of f_1 with respect to θ_3 . $h_{r_1}^0$, \bar{h}_{r_1} , h_2^0 , \bar{h}_2 are

functions with the argument s for all $s \in \mathbb{R}$.

4.4.1 Consistency

Theorem 4.5. *Let Assumptions 4.1-4.3 hold, and let $f(z_{i,t}; \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as Model (4.4.1) and satisfy Assumption HH. Moreover, if*

- (a) $\nu_i(\sqrt{T})$ and $\dot{\nu}_i(\sqrt{T})$ are bounded away from zero as $T \rightarrow \infty$, for $i = 1, 2$,
- (b) $\nu_3 = \nu_2$, $\frac{\dot{\nu}_2}{\nu_1} = o_p(1)$, $\frac{\nu_1 \dot{\nu}_1}{\nu_1^2} = o_p(1)$, $\frac{\dot{\nu}_2}{\nu_1 \nu_2 T} = o_p(1)$ and $\frac{\dot{\nu}_1}{\nu_1 \nu_2 T} = o_p(1)$,
- (c) $[h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]B_{x,i}(r) \neq \left(\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]B_{x,i}(s)ds \right)$ for all $(\theta'_1, \theta'_2)' \neq (\theta'_{1,0}, \theta'_{2,0})'$ and some $r \in [0, 1]$ almost surely,
- (d) $[\bar{h}_{r_1}^B - h_{r_1}^{B,0} + \bar{h}_2^B - h_2^{B,0}]B_{x,i}(r) \neq \left(\int_0^1 [\bar{h}_{r_1}^B - h_{r_1}^{B,0} + \bar{h}_2^B - h_2^{B,0}]B_{x,i}(s)ds \right)$ for all $(\theta'_3, \theta'_4)' \neq (\theta'_{3,0}, \theta'_{4,0})'$ and some $r \in [0, 1]$ almost surely,

we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$R^*(\theta_1, \theta_2, \theta_{1,0}, \theta_{2,0}) =$$

$$\mathbf{E} \left[\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]^2 B_{x,i}^2(r) dr - \left(\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]B_{x,i}(s)ds \right)^2 \right],$$

with $M_{1,n,T} = n(T\nu_1(\sqrt{T}))^2$, and

$$R^*(\theta_{1,0}, \theta_{2,0}, \theta_3, \theta_4, \theta_{3,0}, \theta_{4,0}) =$$

$$\mathbf{E} \left[\int_0^1 [\bar{h}_{r_1}^B - h_{r_1}^{B,0} + \bar{h}_2^B - h_2^{B,0}]B_{x,i}(s)ds - \left(\int_0^1 [\bar{h}_{r_1}^B - h_{r_1}^{B,0} + \bar{h}_2^B - h_2^{B,0}]B_{x,i}(r)dr \right)^2 \right],$$

with $M_{2,n,T} = n(T\nu_2(\sqrt{T}))^2$. Moreover,

$$(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' \xrightarrow{P} (\theta'_{1,0}, \theta'_{2,0})',$$

at rate $(T\nu_1)^{-1}$, and

$$(\hat{\theta}'_{3,n,T}, \hat{\theta}'_{4,n,T})' \xrightarrow{P} (\theta'_{3,0}, \theta'_{4,0})',$$

at rate $(T\nu_2)^{-1}$.

Theorem 4.6. *Let Assumptions 4.1-4.3 hold, and let $f(z_{i,t}, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as Model (4.4.1) and satisfy Assumption HI. Moreover, if*

$$(a) \frac{\nu_1}{\nu_1 T^{3/4}} \text{ is } o_p(1),$$

$$(b) [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]B_{x,i}(r) \neq \left(\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]B_{x,i}(s)ds \right) \\ \text{for all } (\theta'_1, \theta'_2)' \neq (\theta'_{1,0}, \theta'_{2,0})' \text{ and some } r \in [0, 1] \text{ almost surely,}$$

$$(c) \int_{-\infty}^{\infty} [\bar{r}_1 - r_1^0 + \bar{f}_2 - f_2^0]^2 ds > 0 \text{ for all } (\theta'_3, \theta'_4)' \neq (\theta'_{3,0}, \theta'_{4,0})',$$

we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$R^*(\theta_1, \theta_2, \theta_{1,0}, \theta_{2,0}) =$$

$$\mathbf{E} \left[\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]^2 B_{x,i}^2(r) dr - \left(\int_0^1 [h_1(B_{z,i}; \theta_1, \theta_2) - h_1(B_{z,i}; \theta_{1,0}, \theta_{2,0})]B_{x,i}(s)ds \right)^2 \right],$$

with $M_{1,n,T} = n(T\nu_1(\sqrt{T}))^2$, and

$$R^*(\theta_{1,0}, \theta_{2,0}, \theta_3, \theta_4, \theta_{3,0}, \theta_{4,0}) = \int_{-\infty}^{\infty} [\bar{r}_1 - r_1^0 + \bar{f}_2 - f_2^0]^2 ds^2 \mathbf{E} \left[\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right],$$

with $M_{2,n,T} = n(T)^{3/2}$. Moreover,

$$(\hat{\theta}'_{1,n,T}, \hat{\theta}'_{2,n,T})' \xrightarrow{P} (\theta'_{1,0}, \theta'_{2,0})',$$

at rate $(T\nu_1)^{-1}$, and

$$(\hat{\theta}'_{3,n,T}, \hat{\theta}'_{4,n,T})' \xrightarrow{P} (\theta'_{3,0}, \theta'_{4,0})',$$

at rate $(\sqrt{n}T^{3/4})^{-1}$. In particular,

Remark 4.3.

- (1) Similar to the conclusion on the Pooled NLS estimator, by reiterating the arguments, the results can be generalized to the K -stage. If the nonlinear function is H -regular, the identification conditions are stronger than that required by the Pooled NLS estimator.
- (2) Comparing the above results with those obtained by Theorems 3.5 and 3.6, the only difference happens when there are H -regular components appearing in the model specification.

4.4.2 Asymptotic Normality

the asymptotic distributions of the Pooled NLSDV estimator for the multiple strong co-summability are presented in the following theorems.

Theorem 4.7. *Let Assumptions 4.1' and 4.2-4.3 hold, and let $f(z_{i,t}; \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as for Model (4.4.1) and satisfy Assumption HH. Moreover, if*

(a) *All assumptions in Theorem 4.5 are satisfied,*

(b) $\frac{\dot{\nu}_2}{\dot{\nu}_1} \xrightarrow{P} 0, \frac{\ddot{\nu}_2}{\dot{\nu}_1} \xrightarrow{P} 0, \ddot{\nu}_3 = \ddot{\nu}_2$ *and* $\dot{\nu}_3 = \dot{\nu}_2,$

(c) $\int_0^1 \mathbf{V}_H \mathbf{V}_H' dr > 0,$ *a.s., where*

$$\mathbf{V}_H = \begin{bmatrix} \dot{h}_{1,1,0} B_{x,i} - \int_0^1 \dot{h}_{1,1,0} B_{x,i} ds \\ \dot{h}_{1,2,0} B_{x,i} - \int_0^1 \dot{h}_{1,2,0} B_{x,i} ds \\ (\dot{h}_{r1,3,0} + \dot{h}_{2,3,0}) B_{x,i} - \int_0^1 [\dot{h}_{r1,3,0} + \dot{h}_{2,3,0}] B_{x,i} ds \\ \dot{h}_{2,4,0} B_{x,i} - \int_0^1 \dot{h}_{2,4,0} B_{x,i} ds \end{bmatrix},$$

for all $\delta > 0$, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\sqrt{n} TC_T(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{d} \mathbf{N}(0, \Sigma_{HH}^{-1}),$$

where

$$C_T = \begin{bmatrix} \dot{\nu}_1 & 0 & 0 & 0 \\ 0 & \dot{\nu}_1 & 0 & 0 \\ 0 & 0 & \dot{\nu}_2 & 0 \\ 0 & 0 & 0 & \dot{\nu}_2 \end{bmatrix},$$

and

$$\Sigma_{HH} = \mathbf{E} \left\{ \int_0^1 \mathbf{V}_H \mathbf{V}_H' dr \right\}.$$

Condition (b) in Theorem 4.7 is needed for the derivation because we consider a general class of functions in the theorem. Hence, it is very difficult to come up with an example that fails the condition if not impossible. However, in practice, most popular H – regular functions such as the polynomial functions satisfy this condition.

Theorem 4.8. *Let Assumptions 4.1-4.3 hold, and let $f(z_{i,t}; \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be defined as for Model (4.4.1) and satisfy Assumption HI. Moreover, if*

(a) *All assumptions in Theorem 4.6 are satisfied,*

(b) $\int_0^1 \mathbf{V}_I \mathbf{V}_I' dr > 0$, a.s., *where*

$$\mathbf{V}_I = \begin{pmatrix} \dot{h}_{1,1,0} B_{x,i} - \int_0^1 \dot{h}_{1,1,0} B_{x,i} ds \\ \dot{h}_{1,2,0} B_{x,i} - \int_0^1 \dot{h}_{1,2,0} B_{x,i} ds \end{pmatrix},$$

and

$$\int_{-\infty}^{\infty} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix}' ds > 0,$$

we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:

$$\sqrt{n}T\dot{\nu}_1 \begin{pmatrix} \hat{\theta}_{1,n,T} - \theta_{1,0} \\ \hat{\theta}_{2,n,T} - \theta_{2,0} \end{pmatrix} \xrightarrow{d} \mathbf{N}(0, \Sigma_{HI,1}^{-1}),$$

where

$$\Sigma_{HI,1} = \mathbf{E} \left[\int_0^1 \mathbf{V}_I \mathbf{V}_I' dr \right].$$

Moreover,

$$\sqrt{n}T^{3/4} \begin{pmatrix} \hat{\theta}_{3,n,T} - \theta_{3,0} \\ \hat{\theta}_{4,n,T} - \theta_{4,0} \end{pmatrix} \xrightarrow{d} \mathbf{N}(0, \Sigma_{HI,2}^{-1}),$$

where

$$\Sigma_{HI,2} = \int_{-\infty}^{\infty} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix} \begin{pmatrix} [\dot{r}_{1,3,0} + \dot{h}_{2,3,0}] \\ \dot{h}_{2,4,0} \end{pmatrix}' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2 dL_{z,i}(r, 0) \right].$$

4.4.3 Estimation of the Covariance Matrix

We define the estimator $\hat{\Sigma}_{n,T}$ for the covariance matrices Σ_{HH} and Σ_{HI} as (4.3.2) where.

$$\Sigma_{HI} = \begin{bmatrix} \Sigma_{HI,1} & 0 \\ 0 & \Sigma_{HI,2} \end{bmatrix}. \quad (4.4.2)$$

Then, the following corollaries show the consistency of $\hat{\Sigma}_{n,T}$ for multiple strong co-summable regression models.

Corollary 4.3. *Let the assumptions of Theorem 4.7 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma_{HH},$$

where $C_{n,T} = \sqrt{n}TC_T$ with C_T and Σ_{HH} being defined in Theorem 4.7.

Corollary 4.4. *Let the assumptions of Theorem 4.8 hold. We then have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$:*

$$\hat{\Sigma}_{n,T} \xrightarrow{p} \Sigma_{HI},$$

where $C_{n,T} = \sqrt{n}C_T$ with

$$C_T = \begin{bmatrix} T\dot{\nu}_1 & 0 & 0 & 0 \\ 0 & T\dot{\nu}_1 & 0 & 0 \\ 0 & 0 & \sqrt[3]{T} & 0 \\ 0 & 0 & 0 & \sqrt[3]{T} \end{bmatrix},$$

and Σ_{HI} being defined in (4.4.2).

4.5 Smooth Transition Co-summable Regressions

In this section, we derive the asymptotic properties of the Pooled NLSDV estimator for the multiple strong co-summable regression defined by the Smooth Transition Regression model with fixed effects:

$$y_{i,t} = \delta_i + \{\alpha + \beta f(z_{i,t}; \gamma, c)\}x_{i,t} + u_{i,t}, \quad (4.5.1)$$

where $x_{i,t}$ and $z_{i,t}$ are the $S(1)$ strictly exogenous variables, and $u_{i,t}$ is a martingale difference process. Furthermore, $z_{i,t}$ is named indicator or transition variable which determines the prevailing regime at time t . The transition function $f(z_{i,t}; \gamma, c)$ in (4.5.1) is a continuous function that is bounded between

0 and 1. In this chapter, $f(z_{i,t}; \gamma, c)$ is chosen to be the logistic function considered by Granger and Teräsvirta (1993):

$$f(z_{i,t}; \gamma, c) = \frac{1}{1 + \exp(-\gamma(z_{i,t} - c))}, \quad (4.5.2)$$

with $\gamma > 0$, which is an identifying restriction. Based on the discussion in Section 6 of Chapter 3, Model (4.5.1) can be reparametrized as:

$$y_{i,t} = \delta_i + f_1(x_{i,t}; \alpha) + f_2(x_{i,t}; \beta)f_3(z_{i,t}) + f_2(x_{i,t}; \beta)f_4(z_{i,t}; \gamma, c) + u_{i,t}, \quad (4.5.3)$$

where $f_1(x_{i,t}; \alpha) = \alpha x_{i,t}$ and $f_2(x_{i,t}; \beta) = \beta x_{i,t}$ are *homogeneous functions* of $x_{i,t}$ with order one. $f_3(z_{i,t}) = 1\{z_{i,t} \geq 0\}$ is a *homogeneous function* of $z_{i,t}$ with order zero. Moreover, $f_4(z_{i,t}; \gamma, c) = 1_{\mathbb{R}-} \frac{1}{1+\exp(-\gamma(z_{i,t}-c))} - 1_{\mathbb{R}+} \frac{1}{1+\exp(\gamma(z_{i,t}-c))}$ is an *integrable function* of $z_{i,t}$.

4.5.1 Consistency and Asymptotic Normality

The proofs of the following corollaries are based on the representation (4.5.3) for the consistency of the Pooled NLSDV estimator of the parameters in Model (4.5.1).

Corollary 4.5. *Let Assumptions 4.1-4.3 hold respectively for $x_{i,t}$, $z_{i,t}$, $u_{i,t}$, we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$,*

$$R^*(\theta, \theta_0) = \mathbf{E} \left[\int_0^1 [(\alpha_0 - \alpha) + (\beta_0 - \beta)1\{B_{z,i} > 0\}]^2 B_{x,i}^2(r) dr \right] \\ - \mathbf{E} \left[\int_0^1 [(\alpha_0 - \alpha) + (\beta_0 - \beta)1\{B_{z,i} > 0\}] B_{x,i}(r) dr \right]^2,$$

with $M_{n,T} = nT^2$. Moreover, for the Pooled NLSDV estimators $\hat{\alpha}_{n,T}$ and $\hat{\beta}_{n,T}$:

$$\hat{\alpha}_{n,T} \xrightarrow{P} \alpha_0,$$

$$\hat{\beta}_{n,T} \xrightarrow{P} \beta_0.$$

Corollary 4.6. *Let Assumptions 4.1-4.3 hold respectively for $x_{i,t}$, $z_{i,t}$, $u_{i,t}$, we have, as $T \rightarrow \infty$*

followed by $n \rightarrow \infty$,

$$R^*(\theta, \theta_0) = \int_{-\infty}^{\infty} \beta_0^2 [1\{s \leq 0\} \frac{1}{1 + \exp(-\gamma_0(s - c_0))} - 1\{s \geq 0\} \frac{1}{1 + \exp(\gamma_0(s - c_0))} \\ - \left(1\{s \leq 0\} \frac{1}{1 + \exp(-\gamma(s - c))} - 1\{s \geq 0\} \frac{1}{1 + \exp(\gamma(s - c))} \right)]^2 ds \mathbf{E} \left[\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right],$$

with $M_{n,T} = nT^{3/2}$. Moreover, for the Pooled NLSDV estimators $\hat{\gamma}_{n,T}$ and $\hat{c}_{n,T}$:

$$\hat{\gamma}_{n,T} \xrightarrow{p} \gamma_0,$$

$$\hat{c}_{n,T} \xrightarrow{p} c_0.$$

The strict inequalities appearing in Corollaries 4.5 and 4.5 are required by Condition (b) of Lemma 3.3 to identify the parameters. Furthermore, they are no more than special cases of what are required by Theorems 4.7 and 4.8. The following corollary shows the asymptotic distribution of the Pooled NLSDV estimator:

Corollary 4.7. *Let Assumptions 4.1' and 4.2-4.3 hold respectively for $x_{i,t}$, $z_{i,t}$ and $u_{i,t}$, we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$,*

$$(a) \sqrt{n}T \left[\begin{pmatrix} \hat{\alpha}_{n,T} \\ \hat{\beta}_{n,T} \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right] \xrightarrow{d} N(0, \Sigma_{\alpha_0, \beta_0}^{-1}),$$

$$(b) \sqrt{n} \sqrt[3]{T} \left[\begin{pmatrix} \hat{\gamma}_{n,T} \\ \hat{c}_{n,T} \end{pmatrix} - \begin{pmatrix} \gamma_0 \\ c_0 \end{pmatrix} \right] \xrightarrow{d} N(0, \Sigma_{\gamma_0, c_0}^{-1}),$$

where

$$\Sigma_{\alpha_0, \beta_0} = \mathbf{E} \left[\int_0^1 \begin{bmatrix} B_{x,i} - \int_0^1 B_{x,i} ds \\ B_{x,i} 1\{B_{z,i} > 0\} - \int_0^1 B_{x,i} 1\{B_{z,i} > 0\} ds \end{bmatrix} \begin{bmatrix} B_{x,i} - \int_0^1 B_{x,i} ds \\ B_{x,i} 1\{B_{z,i} > 0\} - \int_0^1 B_{x,i} 1\{B_{z,i} > 0\} ds \end{bmatrix}' dr \right],$$

and

$$\Sigma_{\gamma_0, c_0} = \int_{-\infty}^{\infty} \beta_0^2 \begin{bmatrix} \dot{f}_{\gamma_0}(s; \gamma_0, c_0) \\ \dot{f}_{c_0}(s; \gamma_0, c_0) \end{bmatrix} \begin{bmatrix} \dot{f}_{\gamma_0}(s; \gamma_0, c_0) \\ \dot{f}_{c_0}(s; \gamma_0, c_0) \end{bmatrix}' ds \mathbf{E} \left[\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right],$$

with \dot{f}_{γ_0} and \dot{f}_{c_0} being the first order partial derivatives of the transition function with respect to γ

and c .

Both \dot{f}_{γ_0} and \dot{f}_{c_0} satisfy Assumption I. The result (a) only holds for strict exogenous $x_{i,t}$ and $z_{i,t}$. Otherwise, $\hat{\alpha}_{n,T}$ and $\hat{\beta}_{n,T}$ are still consistent but the modification is necessary to guarantee $\sqrt{n}T$ rate of consistency. However, $\sqrt{n}^{3/4}\sqrt{T}$ is the rate of consistency of $\hat{\gamma}_{n,T}$ and $\hat{c}_{n,T}$ in the presence of weak exogeneity.

4.5.2 Estimation of the Covariance Matrix

In this section, we consider the following five consistent estimators for the covariance matrix of $\hat{\theta}_{n,T}$ with $\hat{\theta}_{n,T} = (\hat{\alpha}_{n,T}, \hat{\beta}_{n,T}, \hat{\gamma}_{n,T}, \hat{c}_{n,T})'$ in which the second one is based on the two-step Pooled NLS estimator. We define

$$\hat{\Sigma}_{1,n,T} = \hat{\Sigma}_{n,T},$$

where $\hat{\Sigma}_{n,T}$ is defined in (4.3.2), and

$$\hat{\Sigma}_{2,n,T} = C_{n,T}^{-1} \begin{bmatrix} \sum_i \sum_t \bar{x}_i^2 & \sum_i \sum_t \bar{x}_i \bar{x} z_i & 0 & 0 \\ \sum_i \sum_t \bar{x}_i \bar{x} z_i & \sum_i \sum_t \bar{x} z_i^2 & 0 & 0 \\ 0 & 0 & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t}^2} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} \\ 0 & 0 & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{\gamma}_{n,t} \partial \hat{c}_{n,t}} & \frac{\partial^2 \hat{Q}_{n,T}}{\partial \hat{c}_{n,t}^2} \end{bmatrix} C_{n,T}^{-1},$$

where $\bar{x}_i = x_{i,t} - \frac{1}{T} \sum_{t=1}^T x_{i,t}$, $\bar{x} z_i = x_{i,t} 1\{z_{i,t} \geq 0\} - \frac{1}{T} \sum_{t=1}^T x_{i,t} 1\{z_{i,t} \geq 0\}$ and $\hat{Q}_{n,T}$ is the objective function which is evaluated at the estimates, and

$$C_{n,T} = \begin{bmatrix} T\sqrt{n} & 0 & 0 & 0 \\ 0 & T\sqrt{n} & 0 & 0 \\ 0 & 0 & T^{3/4}\sqrt{n} & 0 \\ 0 & 0 & 0 & T^{3/4}\sqrt{n} \end{bmatrix}.$$

Furthermore, we consider three other estimators that have the same asymptotic limit as those of $\hat{\Sigma}_{1,n,T}$ and $\hat{\Sigma}_{2,n,T}$. We define $\hat{\Sigma}_{3,n,T}$ as $\hat{\Sigma}_{1,n,T}$ excluding the correlation terms between $(\alpha, \beta)'$ and $(\gamma, c)'$, $\hat{\Sigma}_{4,n,T}$ as $\hat{\Sigma}_{1,n,T}$ excluding the second order derivatives of the transition function with respect to γ and c , and $\hat{\Sigma}_{5,n,T}$ as $\hat{\Sigma}_{1,n,T}$ excluding both the second derivatives and the cross correlation terms.

Corollary 4.8. *Let Assumptions 4.1', 4.2-4.3 hold respectively for $x_{i,t}$, $z_{i,t}$, $u_{i,t}$ and f_k with $k = 1, 2, 3, 4$. Then, we have, as $T \rightarrow \infty$ followed by $n \rightarrow \infty$,*

$$\hat{\Sigma}_{l,n,T} \xrightarrow{P} \Sigma_{STR},$$

for $l = 1, 2, 3, 4, 5$, where

$$\Sigma_{STR} = \begin{bmatrix} \Sigma_{\alpha_0, \beta_0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\gamma_0, c_0} \end{bmatrix},$$

with \dot{f}_{γ_0} and \dot{f}_{c_0} being the first partial derivative of the transition function with respect to γ and c .

4.6 Monte Carlo Simulations

We perform 5000 simulations with different combinations of $n = 5, 10, 30, 60, 90, 120$ and $T = 30, 60, 90, 120, 150$ to study the finite sample performance of the Pooled NLSDV estimator. We take Model (4.5.1) and set the values of the parameters as $\alpha_0 = -1.5$, $\beta_0 = 1.5$, $\gamma_0 = 1$ and $c_0 = 0.5$. We only consider the homogeneous panel case with exogenous regressors. For the data generating process, we set, for all i ,

$$\Sigma_i = \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ and, } \varphi_{i,t} = \rho_{i,t} \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix},$$

where $\rho_{i,t}$ is chosen from a uniform distribution on the interval $[0.77, 1.98]$. The innovation terms are assumed to follow AR(1) processes with coefficients $\varphi_{i,t}$. By doing so, we make sure that a unit root is excluded and the time variation is not reduced by $\rho_{i,t}$. δ_i s are generated from a Normal distribution with zero mean and standard deviation 2. Finally, $y_{i,t}$ is generated by using the smooth transition co-summable regression model (4.5.1):

$$y_{i,t} = \delta_i + (-1) \times x_{i,t} + 1.5 \times \frac{1}{1 + \exp(-1 \times (z_{i,t} - 0.5))} x_{i,t} + u_{i,t}, \quad (4.6.1)$$

where $u_{i,t}$ is generated by IID standard *Normal* random variables for each unit i . Moreover, $z_{i,t}$ and $x_{i,t}$ are scalars of regressors assumed to be generated by

$$\begin{bmatrix} x_{i,t} \\ z_{i,t} \end{bmatrix} = \begin{bmatrix} x_{i,t-1} \\ z_{i,t-1} \end{bmatrix} + \begin{bmatrix} v_{x,i,t} \\ v_{z,i,t} \end{bmatrix}, \quad (4.6.2)$$

with common initialization at $t = 0$ satisfying

$$(z_{i,0}, x_{i,0})' = (0, 0)'.$$

The innovation term $V_{i,t} = (v_{x,i,t}, v_{z,i,t})'$ is generated by

$$V_{i,t} = \varphi_{i,t} V_{i,t-1} + \eta_{i,t}, \quad (4.6.3)$$

where $\eta_{i,t}$ is generated by IID standard *Normal* random variables for each unit i .

Tables 4.1-4.4 show the mean absolute error(MAE):

$$MAE = \frac{\sum_{j=1}^{ns} |\hat{\theta}_{n,T} - \theta|}{ns},$$

where $\hat{\theta}_{n,T}$ is the Pooled NLSDV estimates, for different n and T dimensions. The results confirm the asymptotic properties obtained in Section 4.5. First, for any given n , the MAE decreases as T increases. Besides, by moving from the left-upper corner to the right-lower corner of the table, the MAE decreases as n and T jointly increase, which indicates the possibly joint consistency of the Pooled NLSDV estimator. Second, comparing the results for different parameters, the MAEs of $\hat{\alpha}_{n,T}$ and $\hat{\beta}_{n,T}$ are smaller than those of $\hat{\gamma}_{n,T}$ and $\hat{c}_{n,T}$ for the same sample size. This is an indication that the speeds of convergence for the former two are faster than those for the later two.

Figures 4.1-4.8 show the estimated density function of the following transformation of the estimator

$$\sqrt{\hat{\Sigma}_{m,n,T}} C_{n,T} (\hat{\theta}_{n,T} - \theta_0), \quad (4.6.4)$$

with $m = 1, 2$. We only present the empirical probability density functions of the Pooled NLSDV estimator for cross-sectionally independent panel co-summability with strictly exogenous regressors

and $n = 10$ because the performance is good given such a small cross-sectional dimension.² Namely, a kernel estimator is applied to the density functions of the 5000 replications. Figures 1, 3, 5 and 7 show that, based on the estimator $\hat{\Sigma}_{1,n,T}$, the estimated probability density function is close to that of the standard normal distribution, with sample size $n = 10$ and $T = 30, 60, 90, 120, 150$. Moreover, we include the results of $\hat{\Sigma}_{2,n,T}$ for comparison. First, the results show that no matter which estimator is chosen for the covariance matrix, the finite sample distribution of (4.6.4) is close to the standard normal distribution. Second, ignoring the second derivative terms does not cause much distortion from the standard normal distribution. Namely, $\hat{\Sigma}_{1,n,T}$ and $\hat{\Sigma}_{4,n,T}$ are preferred in practice while T is small. Furthermore, the remaining tables show the sample covariances for the 5000 replications of the transformed estimates by (4.6.4). They are very close to the ideal value 1 which is another confirmation of the good finite sample performance of the Pooled NLSDV estimator.

In order to check the robustness of the Pooled NLSDV estimator against possible realizations without piecewise linear properties, an additional simulation study is performed based on the setup defined until 4.6.3 but with

$$\varphi_{i,t} = \rho_{i,t} \begin{bmatrix} -0.4 & 0 \\ 0 & -0.4 \end{bmatrix}. \quad (4.6.5)$$

With 5000 simulations, Tables 3.10-3.13 show the MAE the Pooled NLDV estimates for different combinations of $n = 5, 10, 30, 60, 90, 120$ and $T = 30, 60, 90, 120, 150$. The results are in line with the asymptotic properties derived in Section 4.5. As T and n increase, the MAE decreases and converges to zero, which indicates the consistency of the Pooled NLSDV estimator. Figures 4.9-4.12 show the estimated density function of the transformation of the estimator, which is defined by (4.6.4), with $l = 1$ and $n = 10$. All the estimated density functions are reasonably similar to that of the standard *Normal* random variable. As T increases, the similarity increases as shown in the figures. Hence, the performance of the Pooled NLSDV estimator is robust against the piecewise linear or nonlinear patterns observed in the realizations of Model 4.6.1 in finite samples.

4.7 Conclusions

In this chapter, we propose the Pooled NLSDV estimator for cross-sectionally independent nonlinear panel models with additive fixed individual effects. In general, the Pooled NLSDV estimator is defined as the Pooled NLS estimator for an auxiliary regression equation which is generated by the difference

²The results for $n = 30, 60, 90, 120$ are similar and available upon request.

between the original regression equation and the time average equation. This auxiliary equation defines an objective function for estimation, which is different from that defined by the original regression equation, but retains the parameters of interests. Given that the auxiliary regression equation always holds by the mathematical construction, the theoretical analysis of the asymptotic properties of the Pooled NLSDV estimator is therefore based on the auxiliary regression.

In addition, the linear Pooled LSDV estimator is a special case of the Pooled NLSDV estimator. Given that the Pooled LS estimator is a special case of the Pooled NLS estimator, some conclusions obtained in the linear case still hold in the nonlinear case. Accordingly, the Pooled NLSDV estimator is shown to have the same rate of convergence as that of the Pooled NLS estimator when $T \rightarrow \infty$ being followed by $n \rightarrow \infty$. If the nonlinear function is H -regular, the Pooled NLSDV estimator requires a set of conditions stronger than those required by the Pooled NLS estimator in order to identify the parameters. The implication of the identification conditions depends on the exact form of nonlinearity. For example, if the function is linear, the set of conditions are not necessary. Besides, the asymptotic distribution of the Pooled NLSDV estimator generally has a larger variance than that of the Pooled NLS estimator due to the demeaning. On the other hand, if the function is I -regular, the Pooled NLSDV estimator has the same asymptotic properties as the Pooled NLS estimator. This is due to the asymptotic orthogonality between I -regular functions and deterministic components. The simulation study based on the Logistic Smooth Transition Regression specification confirms the above conclusion and shows that the Pooled NLSDV estimator performs well in finite samples. In view of the applications in González et al. (2005) and Fok et al. (2005), the results in this chapter confirm the validity of their methodology if the data used are considered to be $I(1)$.

Finally, a number of concluding remarks can be made regarding the limitations and possible extensions. First, the assumption of cross-sectional independence is not regarded as being in line with reality, in particular, provided the impressive development in the literature on linear factor models. Second, the assumption of exogeneity seems restrictive for many empirical applications. To the author's best knowledge, for polynomial functions, the issue of endogeneity have been well studied by Hong and Wagner (2008) and Hong and Wagner (2011), as well as for H -regular functions in the time series literature such as Saikkonen and Choi (2004), De Jong (2005) and Kasparis (2008). However, in case of the I -regular function, the issue is still far from being completely understood. The most recent progress can be found in Chang and Park (2005). Third, our simulation study indicates that under cross section independence the Pooled NLS estimator is also consistent as T and n jointly diverge to infinity. However, its theoretical derivation requires further research.

4.A Tables

Table 4.1: MAE for Pooled NLSDV: $\hat{\alpha}_{n,T}$

N \ T	30	60	90	120	150
5	0.0319	0.0121	0.0092	0.0056	0.0044
10	0.0165	0.0071	0.0046	0.0034	0.0027
30	0.0081	0.0037	0.0024	0.0018	0.0014
60	0.0055	0.0026	0.0017	0.0012	0.0010
90	0.0045	0.0021	0.0014	0.0010	0.0008
120	0.0039	0.0018	0.0012	0.0009	0.0007

Table 4.2: MAE for Pooled NLSDV: $\hat{\beta}_{n,T}$

N \ T	30	60	90	120	150
5	0.0641	0.0192	0.0140	0.0085	0.0067
10	0.0244	0.0105	0.0067	0.0050	0.0038
30	0.0114	0.0053	0.0035	0.0026	0.0020
60	0.0080	0.0036	0.0024	0.0018	0.0014
90	0.0062	0.0030	0.0019	0.0014	0.0011
120	0.0054	0.0026	0.0017	0.0012	0.0010

Table 4.3: MAE for Pooled NLSDV: $\hat{\gamma}_{n,T}$

N \ T	30	60	90	120	150
5	0.1955	0.0947	0.0540	0.0459	0.0353
10	0.0791	0.0414	0.0303	0.0247	0.0202
30	0.0375	0.0216	0.0157	0.0126	0.0104
60	0.0257	0.0146	0.0108	0.0085	0.0072
90	0.0210	0.0119	0.0089	0.0069	0.0059
120	0.0180	0.0104	0.0075	0.0060	0.0050

Table 4.4: MAE for Pooled NLSDV: $\hat{c}_{n,T}$

N \ T	30	60	90	120	150
5	0.1944	0.0963	0.0694	0.0528	0.0448
10	0.0966	0.0521	0.0377	0.0293	0.0253
30	0.0477	0.0268	0.0192	0.0153	0.0131
60	0.0335	0.0182	0.0130	0.0104	0.0087
90	0.0269	0.0149	0.0107	0.0084	0.0071
120	0.0226	0.0127	0.0093	0.0074	0.0063

Table 4.5: Estimated Variance for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{1,n,T}$

T \ <i>Trans</i>	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.1137	1.0745	1.1856	1.0612
60	0.9815	1.0008	0.9906	0.9888
90	1.0305	1.0212	0.9990	1.0047
120	0.9815	0.9987	1.0247	0.9704
150	0.9842	0.9744	0.9866	1.0042

Table 4.6: Estimated Variance for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{2,n,T}$

T \ <i>Trans</i>	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.1775	1.6092	1.2173	1.1834
60	1.0247	1.3114	1.1107	1.0761
90	1.0749	1.2838	1.1010	1.0695
120	1.0067	1.2389	1.1146	1.0235
150	1.0155	1.1614	1.0588	1.0519

Table 4.7: Estimated Variance for Pooled NLSDV: $n = 10$,
 $\hat{\Sigma}_{3,n,T}$

$T \setminus Trans$	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.2228	1.3311	1.3824	1.2157
60	1.0233	1.1484	1.1069	1.0758
90	1.0742	1.1523	1.0990	1.0690
120	1.0067	1.1289	1.1147	1.0235
150	1.0150	1.0703	1.0582	1.0517

Table 4.8: Estimated Variance for Pooled NLSDV: $n = 10$,
 $\hat{\Sigma}_{4,n,T}$

$T \setminus Trans$	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.0626	1.0233	1.0200	1.0295
60	0.9815	1.0008	0.9945	0.9890
90	1.0305	1.0212	1.0009	1.0053
120	0.9815	0.9987	1.0247	0.9704
150	0.9842	0.9744	0.9872	1.0044

Table 4.9: Estimated Variance for Pooled NLSDV: $n = 10$,
 $\hat{\Sigma}_{5,n,T}$

$T \setminus Trans$	$\hat{\alpha}_{n,t}$	$\hat{\beta}_{n,T}$	$\hat{\gamma}_{n,T}$	$\hat{c}_{n,T}$
30	1.1717	1.2856	1.2173	1.1834
60	1.0233	1.1484	1.1107	1.0761
90	1.0742	1.1523	1.1010	1.0695
120	1.0067	1.1289	1.1146	1.0235
150	1.0150	1.0703	1.0588	1.0519

Table 4.10: MAE for Pooled NLSDV: $\hat{\alpha}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.1763	0.0548	0.0343	0.0226	0.0179
10	0.0748	0.0300	0.0191	0.0137	0.0105
30	0.0357	0.0157	0.0099	0.0073	0.0057
60	0.0243	0.0109	0.0069	0.0050	0.0039
90	0.0197	0.0088	0.0055	0.0040	0.0032
120	0.0170	0.0076	0.0048	0.0035	0.0027

Table 4.11: MAE for Pooled NLSDV: $\hat{\beta}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.4277	0.1105	0.0606	0.0390	0.0329
10	0.1449	0.0526	0.0310	0.0214	0.0162
30	0.0626	0.0252	0.0158	0.0110	0.0084
60	0.0424	0.0174	0.0107	0.0075	0.0057
90	0.0338	0.0138	0.0086	0.0060	0.0047
120	0.0288	0.0120	0.0075	0.0052	0.0040

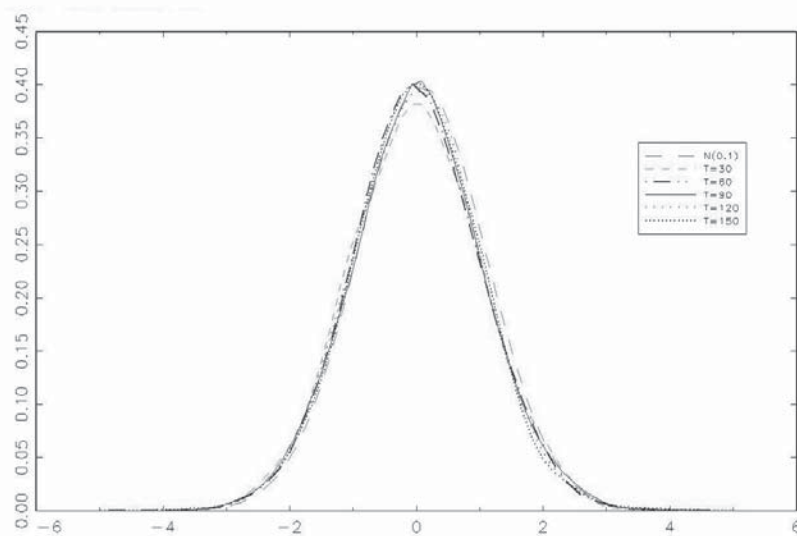
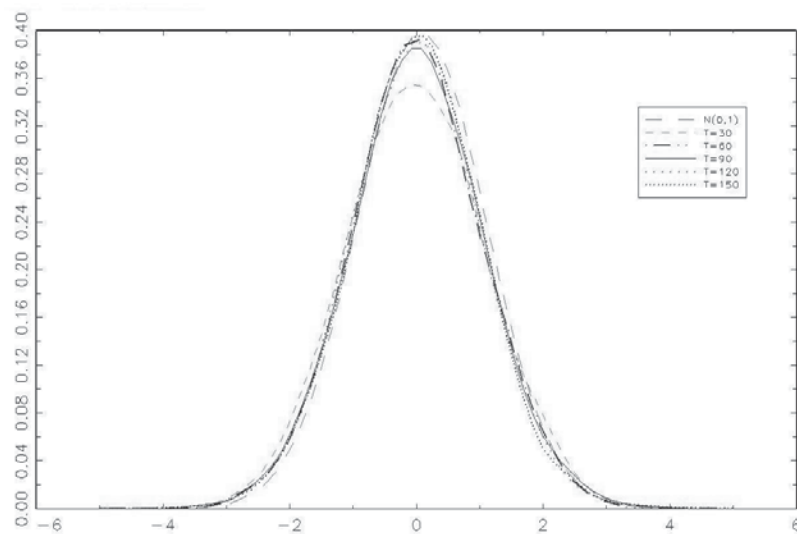
Table 4.12: MAE for Pooled NLSDV: $\hat{\gamma}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.4518	0.1403	0.0888	0.0658	0.0522
10	0.1782	0.0809	0.0516	0.0393	0.0315
30	0.0862	0.0405	0.0279	0.0204	0.0166
60	0.0593	0.0281	0.0189	0.0141	0.0117
90	0.0477	0.0228	0.0152	0.0114	0.0095
120	0.0408	0.0199	0.0133	0.0100	0.0078

Table 4.13: MAE for Pooled NLSDV: $\hat{c}_{n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

N \ T	30	60	90	120	150
5	0.4917	0.1803	0.1123	0.0830	0.0709
10	0.2096	0.0950	0.0628	0.0479	0.0380
30	0.0980	0.0486	0.0335	0.0244	0.0203
60	0.0673	0.0335	0.0228	0.0170	0.0139
90	0.0540	0.0269	0.0184	0.0137	0.0113
120	0.0472	0.0228	0.0158	0.0122	0.0095

4.B Figures

Figure 4.1: Estimated PDF for Pooled NLSDV: $n = 10$, α , $\hat{\Sigma}_{1,n,T}$ Figure 4.2: Estimated PDF for Pooled NLSDV: $n = 10$, α , $\hat{\Sigma}_{2,n,T}$

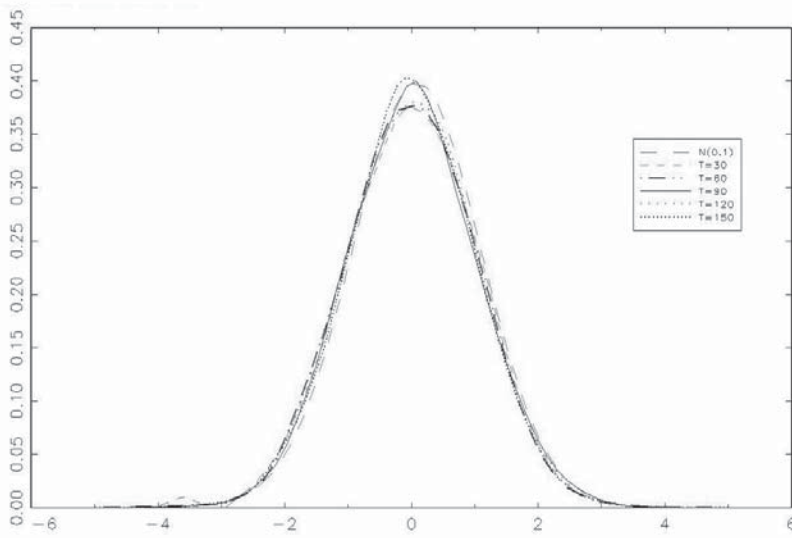


Figure 4.3: Estimated PDF for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{1,n,T}$

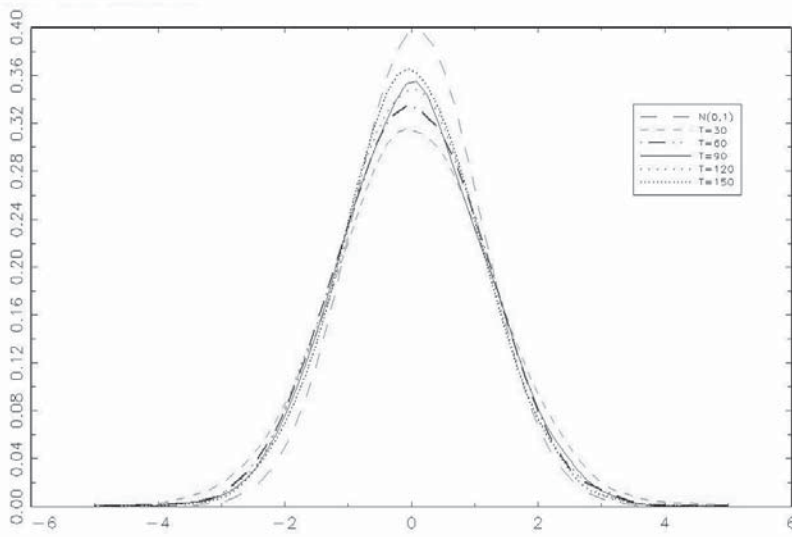


Figure 4.4: Estimated PDF for Pooled NLSDV: $n = 10$, $\hat{\Sigma}_{2,n,T}$

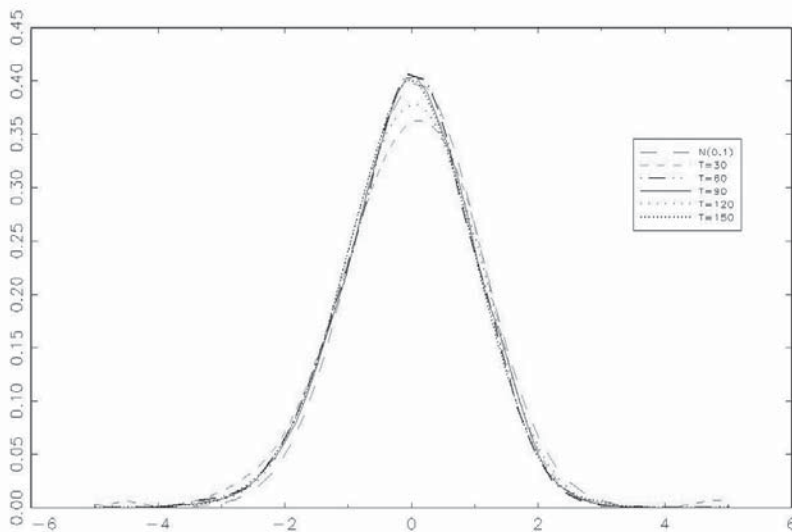


Figure 4.5: Estimated PDF for Pooled NLSDV: $n = 10$, $\gamma, \hat{\Sigma}_{1,n,T}$

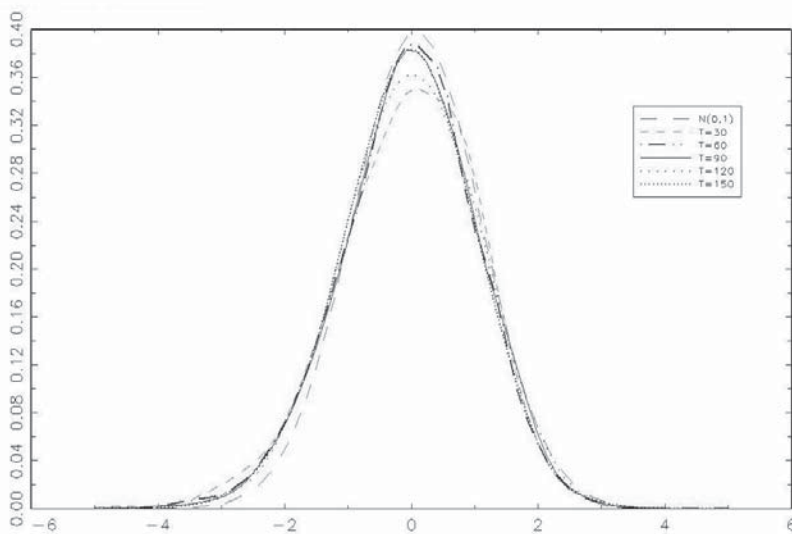


Figure 4.6: Estimated PDF for Pooled NLSDV: $n = 10$, $\gamma, \hat{\Sigma}_{2,n,T}$

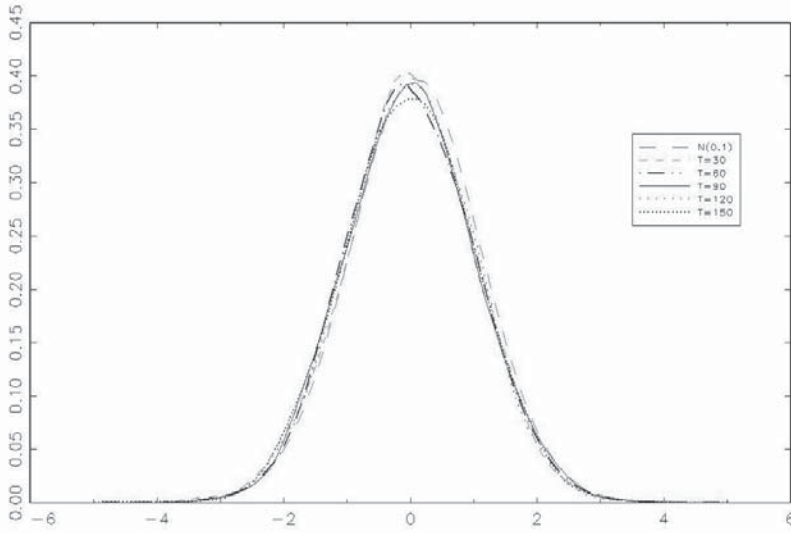


Figure 4.7: Estimated PDF for Pooled NLSDV: $n = 10$, $c, \hat{\Sigma}_{1,n,T}$

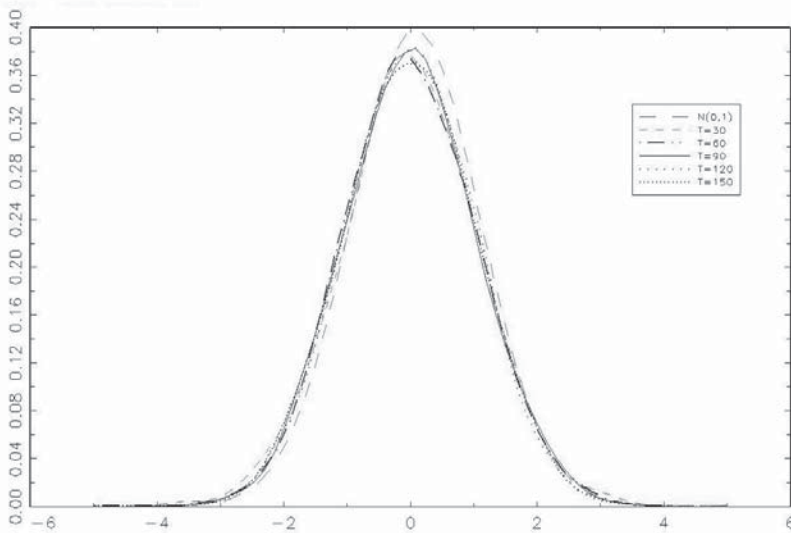


Figure 4.8: Estimated PDF for Pooled NLSDV: $n = 10$, $c, \hat{\Sigma}_{2,n,T}$

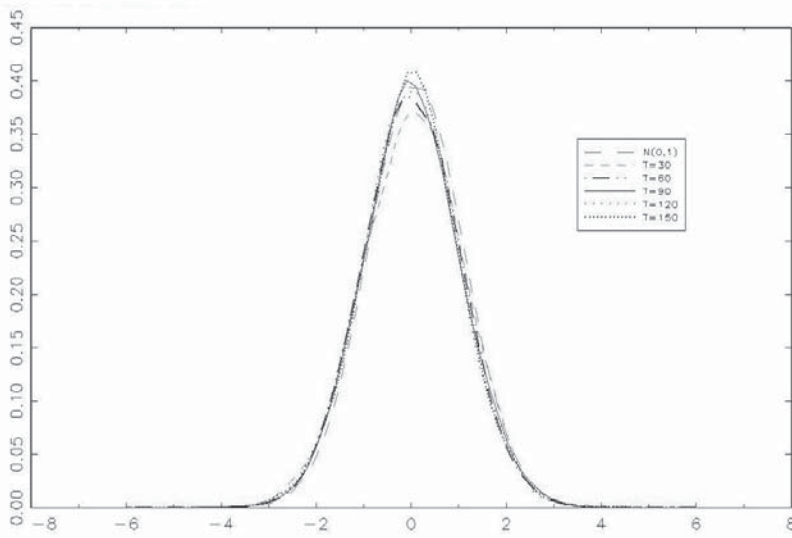


Figure 4.9: Estimated PDF for Pooled NLSDV: $n = 10$, α , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

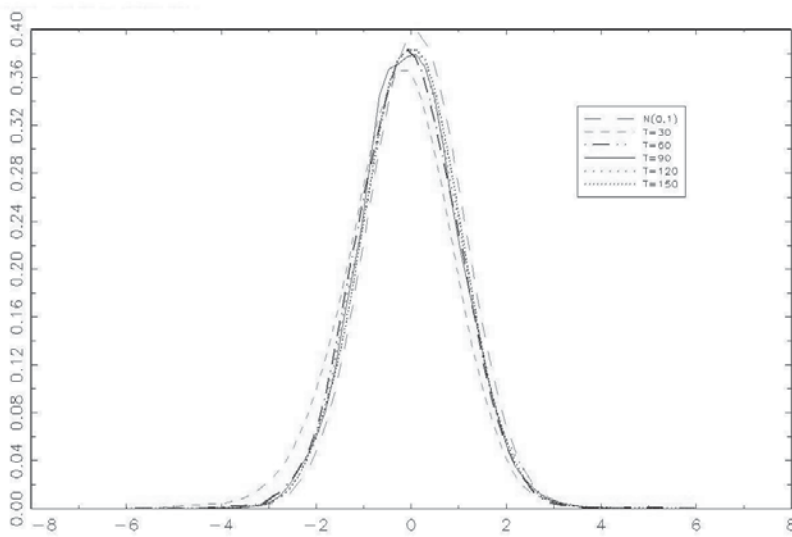


Figure 4.10: Estimated PDF for Pooled NLSDV: $n = 10$, β , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

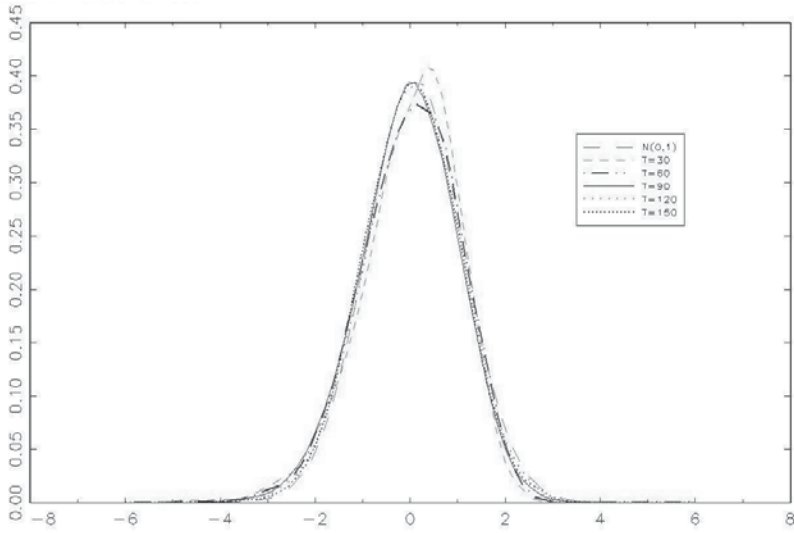


Figure 4.11: Estimated PDF for Pooled NLSDV: $n = 10$, γ , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

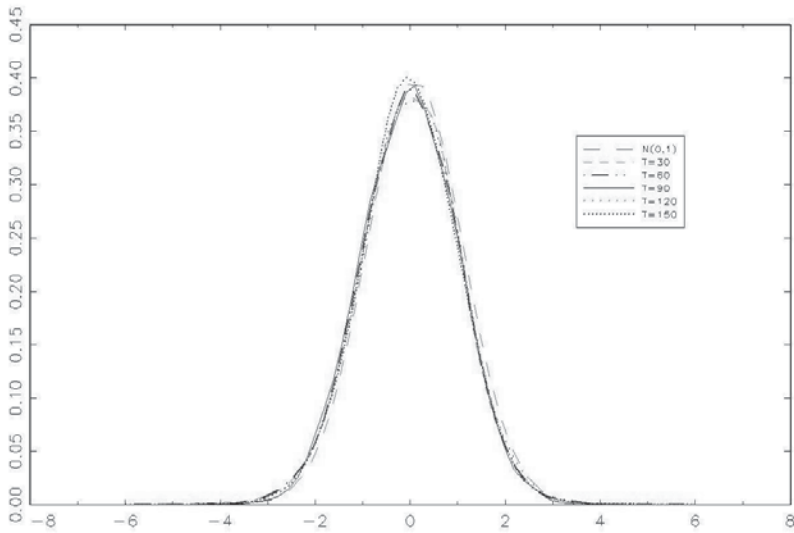


Figure 4.12: Estimated PDF for Pooled NLSDV: $n = 10$, c , $\hat{\Sigma}_{1,n,T}$, $COV(x_{i,t}, z_{i,t}) = 0$

4.C Appendix

4.C.1 Proof of Theorem 4.1

We need to verify the set of sufficient conditions in Lemma 3.3 from Chapter 3. Condition (c) follows from Assumption 4.3. For condition (b),

- As will be shown below, $R^*(\theta, \theta_0)$ is the expectation of nonnegative random variables due to the large n asymptotics. Hence, it is deterministic and nonnegative.
- Continuity follows from uniform convergence. Because $Q_{n,T}$ is continuous, $R_{n,T}^*$ is the difference of two continuous functions and hence it is continuous. Uniform convergence keeps the continuity property.
- The uniqueness of θ_0 is due to condition (b) required by the theorem.

Now, we only need to define $M_{n,T}$, $R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0)$ and $R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0)$, and verify the following uniform convergence

$$M_{n,T}^{-1} R_{n,T,1}^*(\theta, \theta_0) \xrightarrow{P} R^*(\theta, \theta_0), \quad (4.C.1)$$

$$M_{n,T}^{-1} R_{n,T,2}^*(\theta, \theta_0) \xrightarrow{P} 0. \quad (4.C.2)$$

According to (4.2.7) in Section 4.2 and (4.3.1) in Section 4.3, we have

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T \left(g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0) - \frac{1}{T} \sum_{t=1}^T [g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0)] \right)^2,$$

and,

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T \left(g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0) - \frac{1}{T} \sum_{t=1}^T [g(z_{i,t}, x_{i,t}, \theta) - g(z_{i,t}, x_{i,t}, \theta_0)] \right) \left(u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right),$$

where $g(z_{i,t}, x_{i,t}, \theta) = f(z_{i,t}, \theta_0)x_{i,t}$. According to Lemma A6 of Park and Phillips (2001), $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))^2$ and $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))$ are H -regular with asymptotic order $(\nu(\sqrt{T}))^2$ and $\nu(\sqrt{T})$ respectively.

We take $M_{n,T} = nT(\nu(\sqrt{T}))^2$, by Theorem 3.3 of Park and Phillips (2001),

$$\frac{R_{n,T,1}^*(\theta, \theta_0)}{nT(\nu(\sqrt{T}))^2} \xrightarrow{d} \frac{\sum_{i=1}^n \int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 B_{x,i}^2(r) dr - \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)] B_{x,i}(r) dr \right)^2}{n},$$

uniformly in Θ as $T \rightarrow \infty$. Because of the assumption of cross-sectional independence, we have the following for $n \rightarrow \infty$,

$$\frac{R_{n,T,1}^*(\theta, \theta_0)}{nT(\nu(\sqrt{T}))^2} \xrightarrow{a.s.} \mathbf{E} \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)]^2 B_{x,i}^2(r) dr - \left(\int_0^1 [h(B_{z,i}, \theta) - h(B_{z,i}, \theta_0)] B_{x,i}(r) dr \right)^2 \right).$$

Moreover, based on Theorem 3.3 of Park and Phillips (2001), we have

$$R_{n,T,2}^*(\theta, \theta_0) = nO_p(T\nu(\sqrt{T})),$$

as $T \rightarrow \infty$ for fixed n . Hence,

$$\frac{R_{n,T,2}^*(\theta, \theta_0)}{nT(\nu(\sqrt{T}))^2} = \frac{1}{T\nu(\sqrt{T})} O_p(1) \rightarrow 0,$$

uniformly in Θ because $\frac{1}{T\nu(\sqrt{T})} \rightarrow 0$ independent from θ . The uniform convergence follows from the similar arguments in the proof of Lemma 3.1 from Chapter 3 with the locally equicontinuity condition introduced in Definition 3.2 of Park and Phillips (2001). \square

4.C.2 Proof of Theorem 4.2

The arguments are similar to those in the proof of Theorem 3.1 of Chapter 3. The difference is due to $R_{n,T,1}^*(\hat{\theta}_{n,T}, \theta_0)$ and $R_{n,T,2}^*(\hat{\theta}_{n,T}, \theta_0)$. Similar to that of Theorem 4.1, we have

$$R_{n,T,1}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T \left((f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))x_{i,t} - \frac{1}{T} \sum_{t=1}^T [(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))x_{i,t}] \right)^2, \quad (4.C.3)$$

$$R_{n,T,2}^*(\theta, \theta_0) = \sum_{i=1}^n \sum_{t=1}^T \left((f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))x_{i,t} - \frac{1}{T} \sum_{t=1}^T [(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))x_{i,t}] \right) \left(u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right). \quad (4.C.4)$$

Following Lemma A6 of Park and Phillips (2001), $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))^2$ and $(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))$ are I -regular. Hence, using Lemma 3.1 of Chapter 3, taking $k = 2$ and $M_{n,T} = nT^{3/2}$ for $R_{n,T,1}^*(\theta, \theta_0)$:

$$\frac{R_{n,T,1}^*(\theta, \theta_0)}{nT^{3/2}} \xrightarrow{p} \frac{\sum_{i=1}^n \left(\int_{-\infty}^{\infty} [f(s, \theta) - f(s, \theta_0)]^2 ds \right) \left(\int_0^1 B_{x,i}^2(r) dL_{z,i}(r, 0) \right)}{n},$$

uniformly in Θ as $T \rightarrow \infty$. Because of the assumption of cross-sectional independence, we have the convergence result in (4.C.3) as $n \rightarrow \infty$. The extra terms introduced by

$$\frac{1}{T} \sum_{t=1}^T [(f(z_{i,t}, \theta) - f(z_{i,t}, \theta_0))x_{i,t}],$$

disappear here because the presence of $\frac{1}{T}$ is sufficient for the quadratic variation of the I – regular transformations degenerating towards zero. Moreover, based on the second results in Lemma 3.1 of Chapter 3, we have

$$R_{n,T,2}^*(\theta, \theta_0) = nO_p(T^{3/4}),$$

as $T \rightarrow \infty$ for fixed n . Hence,

$$\frac{R_{n,T,2}^*(\theta, \theta_0)}{nT^{3/2}} = \frac{1}{T^{3/4}} O_p(1) \rightarrow 0,$$

uniformly in Θ because $\frac{1}{T^{3/4}} \rightarrow 0$ independent from θ . Finally, (4.C.4) holds. The above pointwise convergence implies uniform convergence in Θ given its compactness and the Lipschitz condition from Definition 3.3 of Park and Phillips (2001). \square

4.C.3 Proof of Theorem 4.3

The proof is in the same spirit as those of Theorem 3.4 of Chapter 3 except for an extra term introduced by the time average. By applying the nonlinear time series asymptotic results in Park and Phillips (2001) with assuming $f(z_{i,t}, \theta)$ being H –regular:

Condition (a): According to the definition of $g(z_{i,t}, x_{i,t}, \theta)$ and $Q_{n,T}(\theta)$, we have

$$\dot{g}(z_{i,t}, x_{i,t}, \theta_0) = \dot{f}(z_{i,t}, \theta_0)x_{i,t}, \quad (4.C.5)$$

$$\begin{aligned} \dot{Q}_{n,T}(\theta_0) &= - \sum_{i=1}^n \sum_{t=1}^T \left[\dot{f}(z_{i,t}, \theta_0)x_{i,t} - \frac{1}{T} \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)x_{i,t} \right] \left[u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right] \\ &\quad - \sum_{i=1}^n \sum_{t=1}^T \left[\dot{f}(z_{i,t}, \theta_0)x_{i,t}u_{i,t} - u_{i,t} \frac{1}{T} \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)x_{i,t} \right], \end{aligned} \quad (4.C.6)$$

and,

$$\begin{aligned} \ddot{Q}_{n,T}^0(\theta_0) &= \sum_{i=1}^n \sum_{t=1}^T [\dot{g}(z_{i,t}, x_{i,t}, \theta_0) - \bar{g}][\dot{g}(z_{i,t}, x_{i,t}, \theta_0) - \bar{g}]', \\ &= \sum_{i=1}^n \sum_{t=1}^T [\dot{f}(z_{i,t}, \theta_0)x_{i,t} - \frac{1}{T} \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)x_{i,t}][\dot{f}(z_{i,t}, \theta_0)x_{i,t} - \frac{1}{T} \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)x_{i,t}]'. \end{aligned} \quad (4.C.7)$$

where $\bar{g} = \frac{1}{T} \sum_{t=1}^T \dot{g}(z_{i,t}, x_{i,t}, \theta_0)$. Now, let $C_{n,T} = \sqrt{nT}\dot{\nu}$. According to Theorem 3.3 from Park and Phillips (2001), we have

$$(T\dot{\nu})^{-1} \sum_{i=1}^n \sum_{t=1}^T \left[\dot{f}(z_{i,t}, \theta_0)x_{i,t}u_{i,t} - u_{i,t} \frac{1}{T} \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0)x_{i,t} \right] \xrightarrow{d} \int_0^1 \dot{h}(B_{z,i}, \theta_0)B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0)B_{x,i}dsdB_{u,i},$$

as $T \rightarrow \infty$. By further applying CLT, for $n \rightarrow \infty$,

$$L_{n,T} \xrightarrow{d} \mathbf{N}(0, \Sigma_H),$$

with

$$\Sigma_H = \mathbf{E} \left[\int_0^1 [\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i}] [\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i}]' dr \right].$$

The above expectation results under the IID property of

$$\int_0^1 [\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i}] [\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i}]' dr,$$

across i , which results from Assumptions 4.1' and 4.2. Next, we establish the convergence of $J_{n,T}$. According to Theorem 3.3 from Park and Phillips (2001) and LLNs, for $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{P} \mathbf{E} \left[\int_0^1 [\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i}] [\dot{h}(B_{z,i}, \theta_0) B_{x,i} - \int_0^1 \dot{h}(B_{z,i}, \theta_0) B_{x,i}]' dr \right].$$

Hence, the results of Theorem 4.3 follow from the convergence of

$$C_{n,T}(\hat{\theta}_{n,T} - \theta_0) = C_{n,T} J_{n,T}^{-1} L_{n,T}.$$

Condition (b): According to Theorem 3.3 from Park and Phillips (2001),

$$(T\ddot{\nu})^{-1} \sum_{t=1}^T \left[\ddot{F}(z_{i,t}, \theta_0) x_{i,t} - \frac{1}{T} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} \right] \left[u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right] = O_p(1).$$

Hence,

$$\begin{aligned} & (T\dot{\nu})^{-1} \sum_{t=1}^T \left[\ddot{F}(z_{i,t}, \theta_0) x_{i,t} - \frac{1}{T} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} \right] \left[u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right] \\ &= (\dot{\nu}^{-1} \ddot{\nu})(\ddot{\nu} T)^{-1} \sum_{t=1}^T \left[\ddot{F}(z_{i,t}, \theta_0) x_{i,t} - \frac{1}{T} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} \right] \left[u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right], \\ &= \|\dot{\nu}^{-1} \ddot{\nu}\| (\ddot{\nu} T)^{-1} \sum_{t=1}^T \left[\ddot{F}(z_{i,t}, \theta_0) x_{i,t} - \frac{1}{T} \sum_{t=1}^T \ddot{F}(z_{i,t}, \theta_0) x_{i,t} \right] \left[u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right], \\ &= O_p(1). \end{aligned}$$

(4.C.8)

The above equality follows from the fact that $\dot{\nu}$ and $\ddot{\nu}$ are strictly positive. Furthermore, $\|\dot{\nu}^{-1}\ddot{\nu}\| < \infty$ is due to condition (a) of Assumption H. In large sample, the above sample moments have expectation zero. Hence,

$$\begin{aligned} R_{n,T} &= C_{n,T}^{-1} \sum_{i=1}^n \sum_{t=1}^T \left[\ddot{G}(z_{i,t}, x_{i,t}, \theta_0) - \frac{1}{T} \sum_{t=1}^T \ddot{G}(z_{i,t}, x_{i,t}, \theta_0) \right] \left[u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t} \right] C_{n,T}^{-1}, \\ &= o_p(1), \end{aligned}$$

for $T \rightarrow \infty$ followed by $n \rightarrow \infty$.

Condition(c): For $T \rightarrow \infty$ followed by $n \rightarrow \infty$,

$$J_{n,T} \xrightarrow{p} J,$$

follows from the proof of Condition (a). Hence, $J > 0$ results from the identification assumption in the theorem.

Condition (d): Given the consistency result from Theorem 4.1 and θ_0 being an interior point, this condition holds trivially.

Condition (e): The time series asymptotic results are the same as those of Theorem 5.2 in Park and Phillips (2001). Hence, we need the following inequalities

$$\begin{aligned} &(\dot{\nu} \otimes \dot{\nu})^{-1} \\ &= (\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu} \nu^{-1} \nu^{-1}, \\ &\leq \|(\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu}\| \|\ddot{\nu}^{-1} \nu^{-1}\|. \end{aligned}$$

$\ddot{\nu}^{-1} \nu^{-1}$ is the convergence rate of the product of f and \ddot{f} . Furthermore, condition (a) of Assumption H implies the boundedness of $\|(\dot{\nu} \otimes \dot{\nu})^{-1} \nu \ddot{\nu}\|$. \square

4.C.4 Proof of Theorem 4.4

The proof mostly follows that of Theorem 3.3 of Chapter 3 except for evaluating the extra terms such as

$$\frac{1}{T^{7/4}} \sum_{i=1}^n \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0) x_{i,t} \sum_{t=1}^T u_{i,t},$$

and

$$\frac{1}{T^{5/2}} \sum_{i=1}^n \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0) x_{i,t} \sum_{t=1}^T \dot{f}(z_{i,t}, \theta_0) x_{i,t}.$$

In fact, the above two sample moments are $o_p(1)$. Hence, the fixed effect is asymptotically orthogonal to the regression components with $I - \text{regular}$ transformation of $I(1)$ processes. \square

4.C.5 Proof of Corollary 4.1

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 4.1, the results in Corollary 4.1 is a natural byproduct of the proof of Theorem 4.3 in view of Condition (e) in Lemma 3.5. \square

4.C.6 Proof of Corollary 4.2

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 4.2, the results in Corollary 4.2 is a natural byproduct of the proof of Theorem 4.4 in view of Condition (e) in Lemma 3.5. \square

4.C.7 Proof of Theorem 4.5

The proof can be simplified by considering the auxiliary equation (4.2.10) and the model (4.4.1). We consider the following elements:

$$f_1(z_{i,t}; \theta_1, \theta_2, \theta_3)x_{i,t} - \frac{1}{T} \sum_{t=1}^T f_1(z_{i,t}; \theta_1, \theta_2, \theta_3)x_{i,t}, \quad (4.C.9)$$

$$f_2(z_{i,t}; \theta_2, \theta_3, \theta_4)\}x_{i,t} - \frac{1}{T} \sum_{t=1}^T f_2(z_{i,t}; \theta_2, \theta_3, \theta_4)\}x_{i,t}, \quad (4.C.10)$$

and

$$u_{i,t} - \frac{1}{T} \sum_{t=1}^T u_{i,t}. \quad (4.C.11)$$

From Lemma 2.2, it follows straightforwardly that,

$$\sum_{t=1}^T f_j x_{i,t},$$

and

$$\sum_{t=1}^T \left[\frac{1}{T} \sum_{t=1}^T f_j x_{i,t} \right],$$

are both $O_p(T^{3/2}\nu_j)$, for $j = 1, 2$. Moreover,

$$\sum_{t=1}^T f_j^2 x_{i,t}^2,$$

and

$$\sum_{t=1}^T \left[\frac{1}{T} \sum_{t=1}^T f_j x_{i,t} \right]^2,$$

are both $O_p(T^2 \nu_j^2)$. Finally, $\sum_{t=1}^T u_{i,t}^2$ and $\sum_{t=1}^T \left[\frac{1}{T} \sum_{t=1}^T u_{i,t} \right]^2$ are both $O_p(T)$.

The results stated in the theorem simply follow from the arguments for Theorem 3.5 by respectively replacing f_1 , f_2 and $u_{i,t}$ with (4.C.9), (4.C.10) and (4.C.11). \square

4.C.8 Proof of Theorem 4.6

The proof can be simplified as that in the proof of Theorem 4.5. The only difference appears when f_2 is assumed to be $I - \text{regular}$. Again, if we consider:

$$f_2(z_{i,t}; \theta_2, \theta_3, \theta_4) \} x_{i,t} - \frac{1}{T} \sum_{t=1}^T f_2(z_{i,t}; \theta_2, \theta_3, \theta_4) \} x_{i,t}. \quad (4.C.12)$$

According to Lemma 3.C.1, $\sum_{t=1}^T f_2^2 x_{i,t}^2$ is $O_p(T^{3/2})$ while $\sum_{t=1}^T \left[\frac{1}{T} \sum_{t=1}^T f_2 x_{i,t} \right]^2$ is $O_p(T)$. Hence, asymptotically, $\sum_{t=1}^T \left[\frac{1}{T} \sum_{t=1}^T f_2 x_{i,t} \right]^2$ disappears from the objective function if it is scaled by $\frac{1}{T^{3/2}}$. The results stated in the theorem simply follow from the arguments for Theorem 3.6 by respectively replacing f_1 , f_2 and $u_{i,t}$ with (4.C.9), (4.C.12) and (4.C.11). \square

4.C.9 Proof of Theorem 4.7

Similar to the proof of Theorem 4.5, the stated results follow directly from the arguments for Theorem 3.7, by respectively replacing f_1 , f_2 and $u_{i,t}$ with (4.C.9), (4.C.10) and (4.C.11). \square

4.C.10 Proof of Theorem 4.8

Similar to the proof of Theorem 4.6, the stated results follow directly from the arguments for Theorem 3.8 if f_1 , f_2 and $u_{i,t}$ are respectively replaced by (4.C.9), (4.C.12) and (4.C.11). \square

4.C.11 Proof of Corollary 4.3

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 4.5, the results in Corollary 4.3 is a natural byproduct of the proof of Theorem 4.7 in view of Condition (e) in Lemma 3.5. \square

4.C.12 Proof of Corollary 4.4

Given the consistency results of $\hat{\theta}_{n,T}$ in Theorem 4.6, the results in Corollary 4.4 is a natural byproduct of the proof of Theorem 4.8 in view of Condition (e) in Lemma 3.5. \square

4.C.13 Proof of Corollary 4.5

Given the LSTCR model is a special case of the Assumption HI, this is the natural consequence of Theorem 4.6. □

4.C.14 Proof of Corollary 4.6

The proof is the same as that of Corollary 4.5. □

4.C.15 Proof of Corollary 4.7

A natural consequence of Theorem 4.8. □

4.C.16 Proof of Corollary 4.8

A natural consequence of Corollaries 4.5, 4.6 and 4.7. □

Polynomial Co-summable Panels with Common Factors

5.1 Introduction

In the econometric literature on linear panel data analysis, a number of different approaches have been advanced to analyze cross-sectional dependence. Based on the availability of measures of economic distance, the techniques from spatial data analysis are often adapted, for example L. Anselin (2001) and Pesaran, Schuermann and Weiner (2004). On the other hand, if the cross-sectional dimension (n) is small and the time series dimension (T) is large, the equations from different cross-sectional individuals can be viewed as a system of seemingly unrelated regression equations, and generalized least squares techniques can be applied. However, neither of the previously mentioned techniques are applicable if both n and T are large, which leads to the studies of modeling cross-sectional dependence with unobserved factors. By assuming a fixed number of factors, Pesaran (2006) proposes approximating the factors by cross-sectional sample averages of the observed variables. The parameters of interests can be consistently estimated by using these approximates to partial out the effects of the unobserved factors. Recently, Kapetanios et al. (2011) extends this technique to linear panel cointegration analysis.

In this chapter, we introduce a panel polynomial co-summable regression model with additive $I(1)$ unobserved common factors, and propose a modified Common Correlated Effects Pooled (CCEP) estimator for the estimation in such models. The polynomial co-summable panel regression model is popular in the economic analysis of the so-called environmental Kuznets curve (EKC) hypothesis. The recent application on EKC and related theoretical work in time series regression models can be found in Hong and Wagner (2008) and Hong and Wagner (2011). Given the availability of many well maintained panel data sets, the analysis of this chapter is expected by potential applications. The results in this chapter are different from that in Kapetanios et al. (2011) in two folders. First, in the presence of nonlinear transformations of the $I(1)$ regressors generally there is no analytical expression for the reduced form of the model so that the Kapetanios et al. (2011) approach cannot be directly

adopted. Second, we consider both the case of regressors being linearly cointegrated with the factors and the case of no such linear cointegration relationship. Our heuristic theoretical analysis shows that, if the coefficients are heterogeneous across individuals, the modified CCEP estimator has the rate of consistency \sqrt{n} in spite of the presence of any cross-sectional cointegration and the order of the polynomial functions. The consistency rate of the modified CCEP estimator for the coefficients of higher order (≥ 2) polynomial functions also depends on T if the coefficients are assumed to be homogeneous across individuals. We do not present a complete theoretical derivation in this chapter which is left for further research. The findings that we input are based on some heuristic derivation and simulation studies.

The structure of this chapter is organized as follows. Section 5.2 introduces the polynomial co-summable panel regression model with common factors, and lays out the assumptions. In Section 5.3, the modified version of the CCEP estimator is introduced. Additionally, the asymptotic results are conjectured for both the case with linear cross section cointegrating relationships and the one without. Section 5.4 reports the results of a Monte Carlo simulation study. Some concluding remarks are made in Section 5.5. Section 5.B contains some heuristic mathematical derivations.

A word about the notations. The standard Euclidean norm of a vector or matrix is denoted by $\|\cdot\|$, i.e. $\|x\|^2 = \sum_i (x_i)^2$. Moreover, the identity matrix is denoted by \mathbf{I} . Letters such as p , q , m , l and δ generally denote positive integers and the definitions should be clear under the context. \xrightarrow{p} denotes the convergence in probability.

5.2 The Model and Assumptions

In this chapter, we consider balanced panels with n cross-sectional units and T time-series observation indexed by $i = 1, \dots, n$ and $t = 1, \dots, T$ respectively. For each unit, the observable variables are respectively the $(k+1)$ -dimensional vector $z_{i,t} = (y_{i,t}, x'_{i,t})'$ and the m_d -dimensional vector d_t , where $\{y_{i,t}\}_{t=0}^T$ is a scalar time series and $\{x_{i,t}\}_{t=0}^T$ is a k -vector time series. The DGP for $\{z_{i,t}\}_{t=0}^T$ can be viewed as the DGP in Kapetanios et al. (2011) augmented by an additional set of independent variables which are correlated with $\{x_{i,t}\}_{t=0}^T$ and the common factors, more specifically,

$$y_{i,t} = \alpha'_i d_t + \gamma'_i f_t + \sum_{j=1}^p \beta'_{j,i} x_{j,i,t} + \varepsilon_{i,t}, \quad (5.2.1)$$

$$x_{i,t} = A'_i d_t + \Gamma'_i f_t + v_{i,t}, \quad (5.2.2)$$

where $x_{j,i,t} = (x_{1,i,t}^j, \dots, x_{k,i,t}^j)$ and f_t is an m_f -dimensional vector. In addition, α_i , γ_i , $\beta_{j,i}$, A_i and Γ_i are $m_d \times 1$, $m_f \times 1$, $k \times 1$, $m_d \times k$ and $m_f \times k$ coefficients or factor loading matrices with fixed and bounded components. d_t consists of deterministic components and possibly I(1) components. f_t and $v_{i,t}$ are I(1) processes. $\varepsilon_{i,t}$ is I(0). We assume the number of common factors is constant and finite in the model.

Remarks 5.1:

- (1) The model specification encompasses the situation where different elements of $x_{i,t}$ consist different orders of polynomial functions $(x_{1,i,t}^j, \dots, x_{k,i,t}^j)$ by simply taking the corresponding elements of $\beta_{j,i}$ as zeros.
- (2) Eq. (5.2.1) defines a strong co-summable relationship among $S(p), \dots, S(2)$ variables with I(0) error terms $\beta'_{1,i} v_{i,t} + \varepsilon_{i,t}$. Otherwise, Eq. (5.2.1) defines a co-summable relationship among $S(p), \dots, S(2)$ variables with I(1) error terms, which are cointegrated with d_t and f_t . In our case¹, where $v_{i,t}$ is assumed to be I(1), there is no cross-sectional cointegration present. Moreover, Eq. (5.2.1) defines a co-summable relationship among $S(p), \dots, S(2)$ variables with I(1) error terms, which are further cointegrated with $v_{i,t}$, d_t and f_t .

Combining (5.2.1) and (5.2.2), we have

$$z_{i,t} = \begin{pmatrix} y_{i,t} \\ x_{i,t} \end{pmatrix} = D'_i d_t + C'_i f_t + \sum_{j=2}^p B_{j,i} x_{j,i,t} + u_{i,t}, \quad (5.2.3)$$

where

$$u_{i,t} = \begin{pmatrix} \varepsilon_{i,t} + \beta'_{1,i} v_{i,t} \\ v_{i,t} \end{pmatrix}, \quad (5.2.4)$$

¹The DGP in (5.2.2) is different from that in Kapetanios et al. (2011), where $v_{i,t}$ are assumed to be I(0). We notice that, by assuming $v_{i,t}$ to be I(0), $x_{i,t}$ are cointegrated with d_t and f_t . Besides, $x_{i,t}$ are cointegrated across i . If

$$\alpha'_i - \beta'_{1,i} A'_i = 0,$$

and

$$\gamma'_i - \beta'_{1,i} \Gamma'_i = 0.$$

$$\begin{aligned}
B_{j,i} &= \begin{pmatrix} \beta'_{j,i} \\ 0 \end{pmatrix}, \\
D_i &= (\alpha_i, A_i) \begin{pmatrix} 1 & 0 \\ \beta_{1,i} & \mathbf{I}_k \end{pmatrix}, \\
C_i &= (\gamma_i, \Gamma_i) \begin{pmatrix} 1 & 0 \\ \beta_{1,i} & \mathbf{I}_k \end{pmatrix}.
\end{aligned} \tag{5.2.5}$$

Furthermore, if

$$\beta_{j,i} = \beta_j + \mu_{j,i},$$

then, the cross-section averages of Eq. (5.2.3):

$$\bar{z}_t = \bar{D}'d_t + \bar{C}'f_t + \sum_{j=2}^p B_j \bar{x}_{j,t} + \bar{u}_t, \tag{5.2.6}$$

where

$$\begin{aligned}
\bar{z}_t &= \frac{1}{n} \sum_{i=1}^n z_{i,t}, \\
\bar{x}_{j,t} &= \frac{1}{n} \sum_{i=1}^n x_{j,i,t},
\end{aligned}$$

$$\begin{aligned}
\bar{u}_t &= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^t \varepsilon_{i,t} + \frac{1}{n} \sum_{i=1}^t \beta'_{1,i} v_{i,t} + \frac{1}{n} \sum_{j=2}^p \sum_{i=1}^t \mu_{j,i} x_{j,i,t} \\ \frac{1}{n} \sum_{i=1}^t v_{i,t} \end{pmatrix}, \\
&= \begin{pmatrix} \bar{\varepsilon}_t + \frac{1}{n} \sum_{i=1}^t \beta'_{1,i} v_{i,t} + \frac{1}{n} \sum_{j=2}^p \sum_{i=1}^t \mu_{j,i} x_{j,i,t} \\ \bar{v}_t \end{pmatrix}.
\end{aligned}$$

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i,$$

and

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i.$$

Moreover, we define

$$X_i = (x_{p,i}, \dots, x_{1,i}),$$

where $x_{j,i} = (x_{j,i,1}, \dots, x_{j,i,T})'$ for $j = 1, \dots, p$. Let

$$y_i = (y_{i,1}, \dots, y_{i,T})'.$$

\bar{M} is defined as

$$\bar{M} = \mathbf{I}_T - \bar{H}(\bar{H}'\bar{H})^{-1}\bar{H}',$$

with $\bar{H} = (D, \bar{X}_p, \dots, \bar{X}_2, \bar{Z})$, D , \bar{X}_j and \bar{Z} being the $T \times m_d$, $T \times k$ and $T \times (k+1)$ matrices of observations on d_t , $\bar{x}_{j,t}$ and \bar{z}_t for $j = 1, \dots, p$. Similarly,

$$M_G = \mathbf{I}_T - G(G'G)^{-1}G',$$

where $G = (D, \bar{X}_p, \dots, \bar{X}_2, F)$ with $D = (d_1, \dots, d_T)'$ and $F = (f_1, \dots, f_T)'$.

The following assumptions are made based on Kapetanios et al. (2011) and Hong and Wagner (2011).

Assumption 5.1. (*Common Factors*)

- (a) d_t are partitioned into $(d'_{1t}, d'_{2t})'$, where d'_{1t} is an m_{d1} -dimensional vector of deterministic component and d'_{2t} is an m_{d2} -dimensional vector of $I(1)$ component.
- (b) The $(m_{d2} + m_f) \times 1$ vector of stochastic common effects, $g_t = (d'_{2t}, f'_t)'$, follows the multivariate unit root process

$$\Delta g_t = \zeta_t,$$

where ζ_t is an $(m_{d2} + m_f) \times 1$ vector of $L_{2+\delta}$, $\delta > 0$, stationary near epoch dependent (NED) processes of size $1/2$, on some α -mixing process of size $-(2 + \delta)/\delta$, with at least $(4 + \delta)$ th moment being finite.

- (c) ζ_t is distributed independently of $\varepsilon_{i,s}$ and $v_{i,s}$ for all i , t and s .

Assumption 5.2. (*Coefficients and Factor Loadings*)

- (a) $\beta_{j,i} = \beta_j + \mu_{j,i}$, for $j = 1, \dots, p$, where β_j is a $k \times 1$ vector of constants. $\mu_{j,i}$ are IID across i with finite second moments $\Omega_{j,\mu}$.
- (b) $\alpha_i = \alpha + \eta_{i,a}$, $\gamma_i = \gamma + \eta_{i,r}$, $A_i = A + \eta_{i,A}$ and $\Gamma_i = \Gamma + \eta_{i,\Gamma}$, where $\eta_{i,a}$, $\eta_{i,r}$, $\eta_{i,A}$ and $\eta_{i,\Gamma}$ are respectively IID across i with finite second moments Ω_a , Ω_r , Ω_A and Ω_Γ .

(c) $\mu_{j,i}$, $\eta_{i,a}$, $\eta_{i,r}$, $\eta_{i,A}$ and $\eta_{i,\Gamma}$ are independent of each other and $\varepsilon_{l,s}$, $v_{l,s}$ and ζ_s for any i , l and s .

Assumption 5.3. (*Idiosyncratic Components I*)

(a) For any i , $(\varepsilon_{i,t}, \Delta v'_{i,t})'$ is an $(k+1) \times 1$ vector of $L_{2+\delta}$, $\delta > 0$, stationary near epoche dependent (NED) processes of size $\delta/(\delta-2)$, on some α -mixing process of size $-(2+\delta)/\delta$, with at least $(4+\delta)$ th moments being finite and independent across i .

(b) $\varepsilon_{i,t}$ and $\Delta v_{l,t}$ are independent of each other for any i , l and t .

Assumption 5.4. (*Identification of β_j*)

(a) $(K_T^{-1} X'_i \bar{M} X_i K_T^{-1})^{-1}$ and $(K_T^{-1} X'_i M_G X_i K_T^{-1})^{-1}$ exist for all i and T^2 .

(b) $\sup_i \|K_T^{-1} X'_i \bar{M} X_i K_T^{-1}\|^2$ is bounded.

(c) $\lim_{(n,T) \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (K_T^{-1} X'_i \bar{M} X_i K_T^{-1})^{-1} = \Omega$, which is nonsingular.

where the $p \times k$ by $p \times k$ dimension square matrix is defined as

$$K_T = \text{diag}(K_p, \dots, K_1),$$

where K_j , for $j = 1, \dots, p$, are $k \times k$ diagonal matrices with $T^{(j+1)/2}$ on the diagonal. Ω requires further derivation and calibration for an explicit expression of the moments defined in Assumption 5.1-5.3. Given that our heuristic arguments are not complete yet in this chapter, we simply assume the existence of Ω .

Assumption 5.5. (*Rank Condition of \bar{C}*)

(a) $rk(\bar{C}) = m \leq k+1$, for all n , and as $n \rightarrow \infty$.

Basically, Assumption 5.4 requires that the total number of dependent and independent variables is at least equal to the number of unobservable common factors. This assumption is not likely to be true in reality. Given the heuristic intention of this chapter, it is not very restrictive for understanding the performance of the CCEP estimator under situations with nonlinearity and without cross-sectional cointegration.

²Any natural number of i or T including T close to infinity.

Assumption 5.6. (*Idiosyncratic Components II*)

- (a) For any i , $(\varepsilon_{i,t}, v'_{i,t})'$ is an $(k+1) \times 1$ vector of $L_{2+\delta}$, $\delta > 0$, stationary near epoqe dependent (NED) processes of size $\delta/(\delta-2)$, on some α -mixing process of size $-(2+\delta)/\delta$, with at least $(4+\delta)$ th moments being finite and independent across i .
- (b) $\varepsilon_{i,t}$ and $v_{l,t}$ are independent of each other for any i, l and t .

5.3 The CCEP Estimator and Asymptotic Results

If Assumption 5.4 holds, the idea of the CCEP estimator in Pesaran (2006) is based on the rewritten form of Eq. (5.2.6):

$$f_t = (\bar{C}\bar{C}')^{-1}\bar{C}[\bar{z}_t - \bar{D}'d_t - \sum_{j=2}^p B_j\bar{x}_{j,t} - \bar{u}_t]. \quad (5.3.1)$$

Since

$$\bar{u}_t \xrightarrow{p} 0,$$

and

$$\bar{C} \xrightarrow{p} C,$$

as $n \rightarrow \infty$, for any t ,

$$f_t - (CC')C[\bar{z}_t - \bar{D}'d_t - \sum_{j=2}^p B_j\bar{x}_{j,t}] \xrightarrow{p} 0.$$

If we plug (5.3.1) back into (5.2.1), we obtain

$$y_{i,t} = (\alpha'_i - P_i\bar{D}')d_t + P_i\bar{z}_t - P_i\sum_{j=2}^p B_j\bar{x}_{j,t} + \sum_{j=1}^p \beta'_{j,i}x_{j,i,t} - P_i\bar{u}_t + \varepsilon_{i,t}, \quad (5.3.2)$$

where $P_i = \gamma'_i(\bar{C}\bar{C}')^{-1}\bar{C}$. The above equation simply suggests that $\beta_{j,i}$ or β_j can be estimated by the method of partial regression. Namely, $y_{i,t}$ and $x_{j,i,t}$ are firstly regressed upon d_t , \bar{z}_t and $\bar{x}_{j,t}$, respectively, along t dimension. Then, using the obtained residuals to estimate $\beta_{j,i}$ or β_j . Hence, the CCEP estimator for model (5.2.1) is defined as

$$\hat{b}_{pool} = \left(\sum_{i=1}^n X'_i \bar{M} X_i \right)^{-1} \left(\sum_{i=1}^n X'_i \bar{M} y_i \right). \quad (5.3.3)$$

Based on the heuristic arguments in Section 5.B, we propose the following propositions:

Proposition 5.1. *If Assumptions 5.1-5.5 hold, considering the model defined by (5.2.1) and (5.2.2), the Common Correlated Effects Pooled estimator, \hat{b}_{pool} , defined by (5.3.3), as $(n, T) \rightarrow \infty$,*

$$\sqrt{n}(\hat{b}_{pool} - \beta) \xrightarrow{P} N(0, \Omega_b),$$

where $\beta = (\beta'_p, \dots, \beta'_1)'$ is a $k \times p$ -dimensional vector.

Proposition 5.2. *If Assumptions 5.1, 5.3-5.6 hold, considering the model defined by (5.2.1) and (5.2.2), the Common Correlated Effects Pooled estimator, \hat{b}_{pool} , defined by (5.3.3), as $(n, T) \rightarrow \infty$,*

$$\sqrt{n}(\hat{b}_{pool} - \beta) \xrightarrow{P} N(0, \Sigma_b),$$

where $\beta = (\beta'_p, \dots, \beta'_1)'$ is a $k \times p$ -dimensional vector.

Remarks 5.2:

- (1) The above propositions only suggest the asymptotic normality of the CCEP estimator but not explicitly specify the covariance matrix, which requires further studies. It is worth noticing that the covariance matrix of the asymptotic distribution is different from those obtained by Pesaran (2006) and Kapetanios et al. (2011) for the linear case. Due to the presence of higher order polynomials, the covariance matrix also depends on the nonlinearity.
- (2) No matter $v_{i,t}$ is I(1) or I(0), the rate of convergence for the CCEP estimator for model (5.2.1) and (5.2.2) only depends on n . Even under the presence of nonlinearity, it is \sqrt{n} which is the same as those obtained by Pesaran (2006) and Kapetanios et al. (2011) for the linear case.
- (3) According to the heuristic arguments in Section 5.B, $X'_i \bar{M} X_i \mu_i$, with $\mu_i = (\mu'_{p,i}, \dots, \mu'_{1,i})'$, is critical for the time series dimension analysis. In particular, $X'_i \bar{M} X_i$, which has the same rate K_T^2 of divergence as that of the denominator of the CCEP estimator, determines that the rate of convergence for the CCEP estimator is independent from T . This term disappears if $\beta_{j,i}$, for $j = 1, \dots, p$, is assumed to be homogeneous across i . In this case, the convergence rate of the CCEP estimator of $\beta_{j,i}$ also depends on T .

5.4 Monte Carlo Simulations

In the simulation study, we consider the following polynomial co-summable panel models with unobservable common factors:

$$y_{i,t} = \alpha_i + \gamma_{1,i}f_{1,t} + \gamma_{2,i}f_{2,t} + \beta_{2,i}x_{i,t}^2 + \beta_{1,i}x_{i,t} + \varepsilon_{i,t}, \quad (5.4.1)$$

where $x_{i,t}$ is a scalar regressor, which is assumed to be generated by

$$x_{i,t} = \Gamma_i f_{1,t} + v_{i,t},$$

with

$$\begin{bmatrix} v_{i,t} \\ \varepsilon_{i,t} \end{bmatrix} - \rho \begin{bmatrix} v_{i,t-1} \\ \varepsilon_{i,t-1} \end{bmatrix} = \sum_{s=0}^{\infty} \phi_{i,t-s} \begin{bmatrix} \epsilon_{x,i,t-s} \\ \epsilon_{\varepsilon,i,t-s} \end{bmatrix}. \quad (5.4.2)$$

In addition,

- α_i is generated by IID Normal distribution with mean zero and variance being equal to 1;
- $f_{1,t}$ is generated by a random walk with zero starting value and standard Normal innovation terms;
- $f_{2,t}$ is generated by a random walk with starting value being 5 and $N(0, 6.25)$ innovation terms;
- $\gamma_{1,i}$ is generated by a Uniform $[-1, 0.5]$ random variable;
- $\gamma_{2,i}$ is generated by a Uniform $[-1.5, 2]$ random variable;
- Γ_i is generated by a Uniform $[-0.5, 1.5]$ random variable;
- $\phi_{i,t-s}$ are generated by 2 by 2 diagonal matrices that are varying over s and i ;
- $(\epsilon_{x,i,t-s}, \epsilon_{\varepsilon,i,t-s})$ are generated by independent bivariate Normal random variables with correlation being zero and variance varying over s and i .

We perform 5000 simulations with different combinations of $n = 5, 10, 30, 60, 90, 120$ and $T = 30, 60, 90, 120, 150$ to show the finite sample performance of the CCEP estimator. We take model (5.4.1) and distinguish two main cases:

- (1) Heterogeneous coefficients:

$$\beta_{2,i} = \beta_2 + \mu_{2,i} = -1.5 + \mu_{2,i},$$

and

$$\beta_{1,i} = \beta_1 + \mu_{1,i} = 1.2 + \mu_{1,i},$$

where $\mu_{2,i}$ follows a Normal distribution with mean being zero and standard deviation being 0.05, and $\mu_{1,i}$ follows a Normal distribution with mean being zero and standard deviation being 0.015, independent of each other for any i .

- (2) Homogeneous coefficients:

$$\beta_{2,i} = \beta_2 = -1.5,$$

and

$$\beta_{1,i} = \beta_1 = 1.2,$$

for any i .

Moreover, for each above mentioned situations, we distinguish two different cases based on the value of ρ :

- (1) The regressors $x_{i,t}$ are not cross-sectionally cointegrated, namely, $v_{i,t}$ are I(1) processes with

$$\rho = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

- (2) The regressors $x_{i,t}$ are cross-sectionally cointegrated, namely, $v_{i,t}$ are I(0) processes with

$$\rho = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Tables 5.1-5.8 report the Mean Absolute Error(MAE) which is defined as

$$MAE = \frac{\sum_{j=1}^{ns} |\hat{b}_{pool} - \beta|}{ns},$$

where $\hat{b}_{pool} = (\hat{b}_{2,pool}, \hat{b}_{1,pool})'$ and $\beta = (\beta_2, \beta_1)'$. The results are concluded as follows:

- (1) Going through all tables, for any given T , MAE decreases as n increases. In particular, the results in Table 5.1-5.2 are in line with the conclusion in Proposition 5.1 and 5.2. Namely, in the

presence of heterogeneous coefficients, the CCEP estimator for the polynomial co-summable regression models is \sqrt{n} consistent.

- (2) If we compare Tables 5.1 and 5.2 with Tables 5.3 and 5.4, the time series properties of $v_{i,t}$ have no effect on the rate of consistency. The same conclusion could be drawn if we compare Tables 5.5 and 5.6 with Tables 5.7 and 5.8.
- (3) By comparing Tables 5.1 and 5.3 with Tables 5.5 and 5.7, we observe that, in the absence of $\mu_{2,i}$ and $\mu_{1,i}$, the rate of consistency shown in Tables 5.5 and 5.7 depends on T . As T increases, the MAE decreases, which is not the case in Tables 5.1 and 5.3. These results are in line with our heuristic arguments in Section 5.B and the discussion in Remark 5.2.
- (4) Repeating the same comparison in (3) for β_1 , we observe that the rate of consistency is independent of T . Namely, it is independent of the presence or the absence of $\mu_{2,i}$ and $\mu_{1,i}$. According to our heuristic arguments in Section 5.B, the numerator of $\hat{b}_{1,pool}$ depends on $X_i' \bar{M} X_i \mu_i$ and $X_i' \bar{M} F \gamma_i$. Given F being I(1) and the corresponding part of X_i being also I(1), the absence of μ_i should not change the rate of divergence. However, it does change the asymptotic covariance matrix.

5.5 Conclusion

Kapetanios et al. (2011) adopts an error multifactor structure for modeling linear panel cointegration relationships and extends the CCEP estimator introduced by Pesaran (2006) to the nonstationary situation. In this chapter, we further extend the results of Kapetanios et al. (2011) in three directions. First, nonlinear transformations of the I(1) regressors are introduced into the model. Second, the CCEP estimator is modified to allow additional exogenous regressors that are correlated with the factors but not necessarily have a linear factor structure. Third, the idiosyncratic components of the regressors are allowed to be I(1). The findings of this chapter show that the modified CCEP estimator is still consistent with rate \sqrt{n} and asymptotically normally distributed if the coefficients are heterogeneous across i . The rate of convergence might depend on T if the coefficients are homogeneous. While for the linear part, the rate of convergence is always independent of T no matter the coefficients are homogeneous or not. The results of the simulation study are clearly in accord with the conclusions.

There are several limitations and open issues requiring further research. First, the derivation is heuristic and incomplete. Hence, the set of assumptions used and the proof of both propositions

need to be further refined. Second, due to the lack of complete derivations, no explicit form of the asymptotic covariance matrix has been derived. Hence, no consistent estimator for the covariance matrix is proposed, which is necessary for inference purposes.

5.A Tables

Table 5.1: Heterogeneous Coefficients: CCEP β_2 , I(1) $v_{i,t}$

N \ T	30	60	90	120	150
5	0.0356	0.0317	0.0297	0.0290	0.0282
10	0.0235	0.0215	0.0206	0.0205	0.0201
30	0.0134	0.0131	0.0126	0.0126	0.0122
60	0.0099	0.0096	0.0093	0.0092	0.0093
90	0.0081	0.0079	0.0077	0.0076	0.0078
120	0.0069	0.0068	0.0068	0.0067	0.0067

Table 5.2: Heterogeneous Coefficients: CCEP β_1 , I(1) $v_{i,t}$

N \ T	30	60	90	120	150
5	0.4235	0.4086	0.3943	0.3902	0.3924
10	0.2543	0.2587	0.2748	0.2801	0.2897
30	0.1364	0.1471	0.1569	0.1739	0.1837
60	0.0951	0.1044	0.1162	0.1235	0.1356
90	0.0731	0.0852	0.0928	0.1028	0.1120
120	0.0634	0.0730	0.0821	0.0882	0.0964

Table 5.3: Heterogeneous Coefficients: CCEP β_2 , I(0) $v_{i,t}$

N \ T	30	60	90	120	150
5	0.0438	0.0389	0.0356	0.0359	0.0354
10	0.0257	0.0231	0.0223	0.0221	0.0208
30	0.0137	0.0126	0.0119	0.0117	0.0116
60	0.0092	0.0085	0.0083	0.0083	0.0082
90	0.0075	0.0069	0.0067	0.0066	0.0066
120	0.0064	0.0059	0.0058	0.0058	0.0057

Table 5.4: Heterogeneous Coefficients: CCEP β_1 , $I(0)$ $v_{i,t}$

N \ T	30	60	90	120	150
5	0.4054	0.3585	0.3330	0.3315	0.3291
10	0.2116	0.1933	0.1824	0.1847	0.1790
30	0.1014	0.0890	0.0857	0.0875	0.0836
60	0.0669	0.0585	0.0568	0.0563	0.0552
90	0.0500	0.0464	0.0451	0.0443	0.0442
120	0.0436	0.0395	0.0381	0.0378	0.0376

Table 5.5: Homogeneous Coefficients: CCEP β_2 , $I(1)$ $v_{i,t}$

N \ T	30	60	90	120	150
5	0.0179	0.0112	0.0088	0.0077	0.0068
10	0.0088	0.0055	0.0045	0.0039	0.0037
30	0.0035	0.0021	0.0017	0.0015	0.0013
60	0.0018	0.0011	0.0008	0.0007	0.0006
90	0.0013	0.0007	0.0005	0.0004	0.0004
120	0.0010	0.0005	0.0004	0.0003	0.0003

Table 5.6: Homogeneous Coefficients: CCEP β_1 , $I(1)$ $v_{i,t}$

N \ T	30	60	90	120	150
5	0.2534	0.1956	0.1758	0.1698	0.1607
10	0.1352	0.1048	0.0966	0.0957	0.0920
30	0.0548	0.0427	0.0379	0.0393	0.0355
60	0.0299	0.0219	0.0188	0.0186	0.0174
90	0.0207	0.0145	0.0121	0.0113	0.0108
120	0.0163	0.0105	0.0089	0.0082	0.0076

Table 5.7: Homogeneous Coefficients: CCEP β_2 , I(0) $v_{i,t}$

N \ T	30	60	90	120	150
5	0.0275	0.0192	0.0152	0.0136	0.0124
10	0.0137	0.0091	0.0075	0.0067	0.0060
30	0.0058	0.0037	0.0028	0.0024	0.0022
60	0.0032	0.0022	0.0016	0.0013	0.0011
90	0.0025	0.0014	0.0011	0.0009	0.0008
120	0.0020	0.0012	0.0009	0.0007	0.0006

Table 5.8: Homogeneous Coefficients: CCEP β_1 , I(0) $v_{i,t}$

N \ T	30	60	90	120	150
5	0.2571	0.2041	0.1874	0.1914	0.1956
10	0.1357	0.1060	0.1026	0.0998	0.0995
30	0.0563	0.0430	0.0369	0.0357	0.0340
60	0.0336	0.0247	0.0200	0.0180	0.0167
90	0.0247	0.0175	0.0141	0.0121	0.0111
120	0.0200	0.0137	0.0111	0.0097	0.0086

5.B Appendix

5.B.1 Proof of Proposition 5.1: Heuristic Arguments

Recall the CCEP estimator from (5.3.3):

$$\hat{b}_{pool} = \left(\sum_{i=1}^n X_i' \bar{M} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' \bar{M} y_i \right). \quad (5.B.1)$$

First, consider $\sum_{i=1}^n X_i' \bar{M} X_i$:

To use Lemma 3 in Kapetanios et al. (2011), we need to prove:

$$K_T^{-1} (X_i' \bar{M} X_i - X_i' M_G X_i) K_T^{-1} = O_p\left(\frac{1}{\sqrt{n}}\right), \quad (5.B.2)$$

as $(n, T) \xrightarrow{j} \infty$. To prove this, we consider the following representation:

$$\bar{H} = G\bar{P} + U^*,$$

where

$$\begin{pmatrix} \mathbf{I}_{m_{d2}} & 0 & \bar{D} \\ 0 & \mathbf{I}_{(p-1) \times k} & B \\ 0 & 0 & \bar{C} \end{pmatrix},$$

with $B = (B'_p, \dots, B'_1)'$. Moreover, $U^* = (0, 0, U)$, where 0s are conformable to the dimension of D and $\bar{X}_p, \dots, \bar{X}_2$, and $U = (\bar{u}_1, \dots, \bar{u}_T)'$. Then, the proof of 5.B.2 is the same as that of Lemma 3 in Kapetanios et al. (2011) and requires further detailed results and derivation similar to those presented in Lemma 1 of Kapetanios et al. (2011).

The last step is to prove

$$\frac{1}{n} \sum_{i=1}^n K_T^{-1} X_i' \bar{M} X_i K_T^{-1} = O_p(1),$$

based on Assumptions 5.1-5.5.

Second, consider $X_i' \bar{M} y_i$:

$$X_i' \bar{M} y_i = X_i' \bar{M} X_i [\mu_i + \varepsilon_i + F\gamma_i].$$

Considering,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n K_T^{-1} X_i' \bar{M} X_i \mu_i K_T^{-1} = O_p(1),$$

due to the independence between $X_i' \bar{M} X_i$ and μ_i , the $O_p(1)$ can be obtained by the CLT.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n K_T^{-1} X_i' \bar{M} \varepsilon_i K_T^{-1},$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n K_T^{-1} X_i' \bar{M} F K_T^{-1} \gamma_i,$$

can be $O_p(1)$ or $o_p(1)$ depending on the highest order of the polynomials. Nevertheless, they are dominated by the order of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n K_T^{-1} X_i' \bar{M} X_i \mu_i K_T^{-1}.$$

Hence, if $\mu_i = 0$, the rate of convergence of the CCEP estimator might depends on K_T . The above argument requires more subtle refinement and studies to complete the following:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i' \bar{M} y_i = O_p(1).$$

The derivation can be based on modifying those in Kapetanios et al. (2011) but the presence of requires \bar{X}_j additional different considerations. □

5.B.2 Proof of Proposition 5.2: Heuristic Arguments

Following the proof of Proposition 5.1. □

Conclusion and Further Research

6.1 Summary, Limitations and Further Research

This thesis provides asymptotic results of a set of pooled estimators, including the Pooled Nonlinear Least Squares estimator, the Pooled Nonlinear Least Squares Dummy Variable estimator and the modified Common Correlated Effects Pooled estimator, for nonlinear panel strong co-summable regression models with both large n and large T dimensions. The main results have been obtained for cross-sectionally independent panels, and the results on the modified Common Correlated Effects Pooled estimator consider nonlinear panel models with persistent cross-sectional dependence, which is assumed to follow a common factor structure. In Chapter 2, we provide an overview of the tools that are required for the statistical interpretation and analysis of nonlinear panel strong co-summable models with large n and large T , which include existing asymptotics for nonlinear transformations of $I(1)$ processes, summability and co-summability, and nonlinear asymptotic theory of stationary processes. The assumptions of a compact parameter space and a twice differentiable objective function with respect to the parameter space seem to be indispensable. Furthermore, in contrast to the analysis of linear models, the derivation of nonlinear asymptotics requires separate consideration of the consistency property and the asymptotic normality.

In Chapter 3, we focus on deriving the sequential asymptotics for the Pooled Nonlinear Least Squares estimator for cross-sectionally independent homogenous panels with large n and large T . In the presence of both nonlinearity and nonstationarity, the concepts of simple strong co-summability and multiple strong co-summability are introduced based on the definitions in Rico (2009). The linear static cointegration model is a typical example of simple strong co-summability. On the other hand, the linear cointegration relationship defined by a set of $I(2)$ processes and some $I(1)$ processes can be viewed as a special case of multiple strong co-summability. The estimator is proved to be consistent and asymptotically normally distributed when $T \rightarrow \infty$ followed by $n \rightarrow \infty$. If the nonlinear transformation is I -regular, the rate of convergence is $T^{3/4}\sqrt{n}$ and weak exogeneity is

sufficient for the asymptotic distribution. For the H -regular functions, the rate is $\nu(T)\sqrt{n}$, where $\nu(T)$ is the homogenous order of the function, and strict exogeneity is required for the asymptotic normality to be free of nuisance parameters. In the case of multiple strong co-summability, the derivation of consistency implies a consistent k -stage Pooled NLS estimator which is not necessarily equivalent to the one-shot Pooled NLS estimator and may have a different rate of consistency and a different asymptotic distribution, due to the presence of nonlinear components. In addition, the reparametrization shows that the Logistic Smooth Transition Co-summable Regression model is actually a special case of the multiple strong co-summable regression models. Both theoretical results and simulation studies indicate that the Pooled NLS estimator performs well for such models. There is no any difficulty in identifying the transition parameters. Finally, the simulation results show that, as T increases, the finite sample distribution of the statistic constructed based on consistent estimators of the covariance matrix is very close to the standard Normal distribution.

In Chapter 4, we propose the Pooled Nonlinear Least Squares Dummy Variable estimator for the models in Chapter 3 augmented with additive fixed individual effects. In contrast to being interpreted as demeaning in the linear case, the Pooled NLSDV estimator is defined as the Pooled NLS estimator for an auxiliary regression equation which is defined as the difference between the original regression equation and the time average equation. This auxiliary equation defines a different objective function from the one defined by the original regression equation, but retains the parameters of interest. Besides, the auxiliary regression equation always holds by mathematical construction. Hence, the Pooled NLSDV estimator is actually a Pooled NLS estimator for a different regression model. Accordingly, the Pooled NLSDV estimator is shown to have the same rate of convergence as that of the Pooled NLS estimator as $T \rightarrow \infty$ being followed by $n \rightarrow \infty$. In the case of the nonlinear function being H -regular, the Pooled NLSDV estimator may require a set of identification conditions stronger than what is required by the Pooled NLS estimator. On the other hand, if the function is I -regular, the Pooled NLSDV estimator has the same asymptotic properties as the Pooled NLS estimator. This confirms the conclusion that I -regular functions are asymptotically orthogonal to deterministic components, which is drawn by Chang et al. (2001). The simulation study based on the Logistic Smooth Transition Regression models confirms the applications in González et al. (2005) and Fok et al. (2005), when the data used there are considered to be $I(1)$.

In Chapter 5, we consider a persistent error multifactor structure for panel polynomial co-summable regression models. We further extend the CCEP estimator proposed by Pesaran (2006) to estimate the parameters of interest. We distinguish two main different situations, namely, parameters

being homogenous across i and parameters being heterogeneous across i . For each of the previously mentioned cases, we consider the effects of both the presence and the absence of cross-sectional cointegration among the regressors. Our heuristic analysis and the simulation study suggest that the modified CCEP estimator is still consistent with rate \sqrt{n} and asymptotically normally distributed if the coefficients are heterogeneous across i . In addition, the simulation results indicate that the rate of consistency depends on T if the coefficients are homogeneous.

Throughout the thesis, we have witnessed the interesting outcomes once the elements of non-linearity, nonstationarity and pooling techniques are combined for panel data models. Moreover, it opens more interesting issues and opportunities for further study and research. First, the DGP used in this thesis assumes the regressors to be either weakly or strictly exogenous, which is considered not to hold in many empirical applications in econometrics. According to the author's understanding, the endogeneity issue has already been tackled, for the H – *regular* functions, by De Jong (2002), Saikkonen and Choi (2004) and Kasparis (2008). Hence, the results in this thesis for H – *regular* functions can be possibly extended, with fully modified estimators, to account for endogenous regressors. On the other hand, for the I – *regular* functions, the endogeneity issue is still not being solved because of the complexity caused by the dependence on the local behavior of the processes. A most recent attempt can be found in Chang and Park (2005).

Second, the concepts of summability and co-summability introduced by Rico (2009) and Rico and Gonzalo (2011) motivate careful reconsideration in some concepts used in linear cointegration analysis, for example, the issue of spurious regression and the k – *stage* procedures discussed in Chapter 2. Moreover, the opportunity is clearly open for extending the results in Chapters 3 and 4 to models with many additive nonlinear components.

Third, the simulation results in Chapters 3 and 4 clearly indicate that the results obtained for sequential asymptotics also hold for joint asymptotics, which is at least true for the Logistic Smooth Transition Co-summable regression models. The related theoretical conclusions still need to be proved. The same applies to Chapter 5, where only heuristic arguments are presented.

Finally, the possible extension of the error multifactor structure to and the validity of the CCEP estimator for panel Logistic Smooth Transition Co-summable regression models are still open issues. It seems clear that the CCEP estimation method is not applicable if the idiosyncratic components of the regressors are assumed to be $I(1)$. This is because then $\sum_{t=1}^T [f(z_{i,t}; \gamma, c)x_{i,t}]^2$ is $O_p(T^{3/2})$ which is lower than $\sum_{t=1}^T v_{i,t}^2 = O_p(T^2)$ in case of $v_{i,t}$ being $I(1)$. In view of (5.3.2), $v_{i,t}$ appears in the error term of the auxiliary equation on which the CCEP estimator is based. The nonlinear

part of $f(z_{i,t}; \gamma, c)x_{i,t}$ has a lower order of summability than the error term has, hence, the nonlinear part cannot be identified by the auxiliary equation. However, the issue is not crystal-clear if $v_{i,t}$ is assumed to be $I(0)$. Another related open issue is to verify the validity of the partial regression approach under the context of being nonlinearly into parameters and nonstationarity.

Bibliography

- Balke, N. and Wohar, M. (1998). Nonlinear Dynamics and Covered Interest Rate Parity, *Empirical Economics* **23**: 535–559.
- Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*, John Wiley & Sons.
- Baum, C. F., Barkoulas, J. T. and Caglayan, M. (2001). Nonlinear Adjustment to Purchasing Power Parity in the Post-Bretton Woods Era, *Journal of International Money and Finance* **20**: 379–399.
- Bec, F. and Rahbek, A. (2004). Vector Equilibrium Correction Models with Non-linear Discontinuous Adjustments, *Econometrics Journal* **7**(2): 628–651.
- Chang, Y. and Park, J. Y. (2003). Index Models with Integrated Time Series, *Journal of Econometrics* **114**: 73–106.
- Chang, Y. and Park, J. Y. (2005). Endogeneity in Nonlinear Regression with Integrated Time Series, Department of Economics, Rice University, working paper.
- Chang, Y., Park, J. Y. and Phillips, P. C. (2001). Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors, *Econometrics Journal* **4**: 1–36.
- Choi, I. and Saikkonen, P. (2009). Tests for Nonlinear Cointegration, *Econometric Theory* **25**: 1–28.
- Christopeit, N. (2009). Weak Convergence of Nonlinear Transformations of Integrated Process: the Multivariate Case, *Econometric Theory* **25**: 1180–1207.
- Chung, K. L. and Williams, R. J. (1990). *Introduction to Stochastic Integration*, 2nd edn, Boston, Birkhäuser.

- Corradi, V. (1995). Nonlinear Transformation of Integrated Time Series: A Reconsideration, *Journal of Time Series Analysis* **16**(6): 539–549.
- Davidson, J. (2009). When Is a Time Series $I(0)$?, in J. Castle and N. Shepherd (eds), *The Methodology and Practice of Econometrics, a festschrift for David F. Hendry*, Oxford University Press.
- De Jong, R. M. (2002). Nonlinear Estimators with Integrated Regressors but without Exogeneity, Ohio State University, Working Paper.
- De Jong, R. M. (2004). Addendum to “Asymptotics for Nonlinear Transformations of Integrated Time Series”, *Econometric Theory* **20**: 627–635.
- De Jong, R. M. (2005). Nonlinear Time Series Models and Weakly Dependent Innovations, Ohio State University, Working Paper.
- De Jong, R. M. (2009). Exponentials of Unit Root Processes, Ohio State University, Working Paper.
- De Jong, R. M. and Hu, L. (2010). A Note on Nonlinear Models with Integrated Regressors and Convergence Order Results, *Economics Letters*, forthcoming.
- De Jong, R. M. and Lee, J. (2008). Exponential Functionals of Integrated Processes, *Economics Letters* **2**: 181–184.
- De Jong, R. M. and Wang, C.-H. (2005). Further Results on the Asymptotics for Nonlinear Transformations of Integrated Time Series, *Econometric Theory* **21**: 413–430.
- Domowitz, I. and White, H. (1982). Misspecified Models with Dependent Observations, *Journal of Econometrics* **20**: 35–38.
- Enders, W. and Falk, B. (1998). Threshold-autoregressive, Median-unbiased, and Cointegration Tests of Purchasing Power Parity, *International Journal of Forecasting* **14**: 171–186.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and Error Correction: Representation, Estimation, and Testing, *Econometrica* **55**(2): 251–76.
- Ermini, L. and Granger, C. W. J. (1993). Some Generalizations on the Algebra of $I(1)$ Processes, *Journal of Econometrics* **58**(3): 369–384.
- Escribano, A. and Mira, S. (2001). Nonlinear Error Correction Models.

- Fok, D., Van Dijk, D. and Franses, P. H. (2005). Forecasting Aggregates Using Panels of Nonlinear Time Series, *International Journal of Forecasting* **21**: 785–794.
- Franses, P. H. and McAleer, M. (1996). Testing for Unit Root and Non-linear Transformations, *Journal of Time Series Analysis* **19**(2): 147–164.
- González, A., Teräsvirta, T. and Van Dijk, D. (2005). Panel Smooth Transition Regression Models, University of Technology Sydney, Working Paper.
- Granger, C. and Hallman, J. (1988). The Algebra of I (1), Board of Governors of the Federal Reserve System (U.S.), Finance and Economics Discussion Series 45.
- Granger, C. W. J. and Hallman, J. (1991). Nonlinear Transformation of Integrated Time Series, *Journal of Time Series Analysis* **12**: 207–224.
- Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford, UK.
- Haldrup, N. (1998). An Econometric Analysis of I(2) Variables, *Journal of Economic Surveys* **12**: 595–650.
- Hong, S. H. and Wagner, M. (2008). Nonlinear Cointegration Analysis and the Environmental Kuznets Curve, Institute for Advanced Studies, Vienna, Austria, Working Paper.
- Hong, S. H. and Wagner, M. (2011). Cointegrating Polynomial Regression: Fully Modified OLS Estimation and Inference, Institute for Advanced Studies, Vienna, Austria, Working Paper.
- Ibragimov, R. and Phillips, P. C. (2008). Regression Asymptotics Using Martingale Convergence Methods, *Econometric Theory* **24**: 888–947.
- Jennrich, R. I. (1969). Asymptotic Properties of Non-linear Least Squares Estimation, *Annals of Mathematical Statistics* **40**: 633–643.
- Kapetanios, G., Pesaran, M. H. and Yamagata, T. (2011). Panels with Non-stationary Multifactor Error Structures, *Journal of Econometrics* **160**(2): 326 – 348.
- Kasparis, I. (2008). Detection of Functional Form Misspecification in Cointegrating Relationships, *Econometric Theory* **24**: 1373–1403.

- Kasparis, I. (2011). Functional Form Misspecification in Regressions with a Unit Root, *Econometric Theory* **27**: 285–311.
- L. Anselin (2001). Spatial Econometrics, in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, pp. 310–330.
- Lo, M. and Zivot, E. (2001). Threshold Cointegration and Nonlinear Adjustment to the Law of One Price, *Macroeconomic Dynamics* **5**: 533–576.
- O’Connell, P. G. J. (1998). Market Frictions and Real Exchange Rates, *Journal of International Money and Finance* **17**: 71–95.
- Park, J. Y. (2003). Strong Approximation for Nonlinear Transformations of Integrated Time Series, Department of Economics, Rice University, working paper.
- Park, J. Y. and Phillips, P. C. (1999). Asymptotics for Nonlinear Transformations of Integrated Time Series, *Econometric Theory* **15**: 269–298.
- Park, J. Y. and Phillips, P. C. (2001). Nonlinear Regressions with Integrated Time Series, *Econometrica* **69**: 117–161.
- Pedroni, P. (2001). Fully Modified OLS for Heterogeneous Cointegrated Panels, *Advances in Econometrics* pp. 93–130.
- Pesaran, M. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, *Econometrica* **74**(4): 967–1012.
- Pesaran, M. H., Schuermann, T. and Weiner, S. M. (2004). Modelling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model, *Journal of Business and Economic Statistics* **22**(2): 129–162.
- Phillips, P. C. and Moon, H. R. (1999a). Functional Form Misspecification in Regressions with a Unit Root, Yale University, Working Paper.
- Phillips, P. C. and Moon, H. R. (1999b). Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica* **67**: 1057–1111.
- Phillips, P. C. and Solo, V. (1992). Asymptotics for Linear Processes, *The Annals of Statistics* **20**(2): 971–1001.

- Pötscher, B. M. (2004). Nonlinear Functions and Convergence to Brownian Motion: Beyond the Continuous Mapping Theorem, *Econometric Theory* **20**: 1–22.
- Pötscher, B. M. and Prucha, I. R. (1991a). Basic Structure of the Asymptotic Thoery in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts, *Econometrics Reviews* **10**: 125–216.
- Pötscher, B. M. and Prucha, I. R. (1991b). Basic Structure of the Asymptotic Thoery in Dynamic Nonlinear Econometric Models, Part II: Asymptotic Normality, *Econometrics Reviews* **10**: 253–325.
- Pötscher, B. M. and Prucha, I. R. (1994). Generic Uniform Convergence and Equicontinuity Concepts for Random Functions, *Journal of Econometrics* **60**: 23–63.
- Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*, Springer-verlag, Berlin.
- Revuz, D. and Yor, M. (1994). *Continuous Martingale and Brownian Motion*, 2nd edn, Springer-verlag, New York.
- Rico, V. B. (2009). Summability of Stochastic Processes: A Generalization of Integration and Co-Integration valid for Non-linear Processes, Universidad Carlos III de Madrid, Working Paper.
- Rico, V. B. and Gonzalo, J. (2011). Summability of Stochastic Processes: A Generalization of Integration and Co-Integration valid for Non-linear Processes, Universidad Carlos III de Madrid, Job Market Paper.
- Saikkonen, P. and Choi, I. (2004). Cointegrating Smooth Transition Regressions, *Econometric Theory* **20**: 301–340.
- Taylor, A. M. (2001). Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-reversion Tests of the Law of One Price, *Econometrica* **69**: 473–498.
- White, H. (1980). Nonlinear Regression on Cross-section Data, *Econometrica* **48**: 721–746.
- Wooldridge, J. M. (1994). Estimation and Inference for Dependent Processes, in R. F. Engel and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. IV, Elsevier, pp. 2639–2738.

Nederlandse samenvatting

De analyse van niet-stationaire tijdreeksen is een van de grote onderzoeksgebieden geweest binnen de tijdreeksconometrie in de laatste tientallen jaren. Empirische studies tonen aan dat tijdreeksen voor variabelen als het bruto binnenlands product, wisselkoersen en aandelenprijzen in het algemeen $I(1)$, oftewel geïntegreerd, zijn. Het lange termijn evenwicht tussen zulke variabelen wordt cointegratie genoemd. Recentelijk zijn er veel empirische onderzoeken gedaan over lineaire panel data modellen met geïntegreerde tijdreeks componenten, iets dat vergemakkelijkt is door de constructie en beschikbaarheid van veel belangrijke panel data sets die gegevens bevatten over verschillende individuen of landen. In het algemeen bevatten deze panel data sets zowel een grote cross-sectionele dimensie (n) en een groot aantal tijdreeks observaties (T). Dit soort situaties hebben de recentelijke vooruitgang teweeg gebracht in theoretisch werk over asymptotiek voor grote n en grote T .

Van de andere kant lijkt de economische theorie te veronderstellen dat de verbanden tussen veel macro-economische variabelen waarschijnlijk niet-lineair zijn vanwege economische cycli of veranderingen van beleidsregimes. Een aantal empirische toepassingen van niet-lineaire regressiemodellen, waaronder het Threshold model, het Smooth Transition model en het Markov Switching model, kan men vinden in de literatuur over internationale economische studies. Echter, de analyse van niet-lineaire regressiemodellen slechts gebaseerd op tijdreeksdata lijkt in veel gevallen niet efficiënt genoeg om de niet-lineariteit te kunnen identificeren. Bijvoorbeeld, in het geval van Smooth Transitions modellen bevatten de tijdreeksdata slechts enkele observaties van de overgangsperiode in vergelijking tot het aantal observaties van de verschillende economische regimes. In dit geval kan, door middel van het samenvoegen (poolen) van observaties in panel data sets, meer informatie worden verkregen over de parameters die van belang zijn als wordt aangenomen dat de parameters gelijke waarden hebben voor de verschillende eenheden. Daarom lijkt het type model en schattingsmethode dat gebaseerd

is op het poolen van observaties een aantrekkelijke manier om efficiënt niet-lineaire economische verbanden te kunnen identificeren en modelleren.

In dit proefschrift proberen we de elementen niet-stationariteit, niet-lineariteit, panel data en poolende schattingstechnieken te combineren, en nemen we het initiatief om asymptotiek voor grote n en grote T te bestuderen voor niet-lineaire panel modellen met niet-stationaire regressoren. Dit proefschrift bevat dus asymptotische resultaten voor een aantal gepoolde schatters, waaronder de Pooled Nonlinear Least Squares schatter, de Pooled Nonlinear Least Squares Dummy Variable schatter, en de aangepaste Common Correlated Effects Pooled schatter, voor niet-lineaire sterke co-summable panel regressiemodellen met zowel grote n als grote T dimensies. De belangrijkste resultaten zijn voor panels met cross-sectionele onafhankelijkheid, terwijl de resultaten voor de aangepaste Common Correlated Effects Pooled schatter uitgaan van niet-lineaire panel modellen met persistente cross-sectionele afhankelijkheid, waarvan wordt aangenomen dat die uit een gemeenschappelijke factor structuur volgt. In Hoofdstuk 2 bieden we een overzicht van de instrumenten die nodig zijn voor de statistische interpretatie en analyse van niet-lineaire panel co-summable modellen met grote n en grote T , waaronder bestaande asymptotiek voor niet-lineaire transformaties van $I(1)$ processen, summability en co-summability, en niet-lineaire asymptotische theorie van statistische processen. De aannames van een compacte parameterruimte en een tweemaal differentieerbare doelfunctie lijken onvermijdelijk te zijn. Daarnaast, in tegenstelling tot de analyse van lineaire modellen, vraagt de afleiding van niet-lineaire asymptotiek om een gescheiden aanpak van de eigenschappen van consistentie, door middel van de eigenschappen van de doelfunctie, en asymptotische normaliteit.

In Hoofdstuk 3 richten we ons op het afleiden van sequentiële asymptotiek voor de Pooled Nonlinear Least Squares schatter voor cross-sectioneel onafhankelijke homogene panels met grote n en grote T . In aanwezigheid van zowel niet-lineariteit en niet-stationariteit worden de concepten van enkelvoudige sterke co-summability en meervoudige sterke co-summability geïntroduceerd op basis van de definities in Rico (2009). Het lineaire statische cointegratie model gedefinieerd door een verzameling van $I(2)$ processen en enige $I(1)$ processen kan worden gezien als een speciaal geval van meervoudige sterke co-summability. De schatter wordt bewezen consistent en asymptotisch normaal te zijn als $T \rightarrow \infty$ gevolgd door $n \rightarrow \infty$. Als de niet-lineaire transformatie I - *regular* is, is de convergentiesnelheid $T^{3/4}\sqrt{n}$, en is zwakke exogeniteit voldoende voor de asymptotische verdeling. For H - *regular* functies is de snelheid $\nu(T)\sqrt{n}$, waar $\nu(T)$ de homogene orde van de functie is, en is strikte exogeniteit noodzakelijk om geen onbekende parameters in de asymptotische normale verdeling te hebben. In het geval van meervoudige sterke co-summability impliceert de afleiding van

consistentie een consistente k -steps Pooled NLS schatter die niet noodzakelijkerwijs equivalent is aan de 1-stap Pooled NLS schatter en die een andere convergentiesnelheid en limietverdeling kan hebben, door de aanwezigheid van niet-lineaire componenten. Daarnaast toont de herparametrisering aan dat het Logistic Smooth Transition Co-summable regressiemodel eigenlijk een speciaal geval is van het meervoudige sterke co-summable regressiemodel. Zowel theoretische als simulatieresultaten tonen aan dat de Pooled NLS schatter goed werkt voor zulke modellen. Er zijn geen problemen om de overgangsparemeters te identificeren. Ten slotte tonen de simulatieresultaten dat, met een groeiende T , de verdeling van de statistiek gebaseerd op de consistente schatters van de covariantie matrix, heel dicht bij de standaard normale verdeling ligt.

In Hoofdstuk 4 introduceren we de Pooled Nonlinear Least Squares Dummy Variables schatter voor de modellen in Hoofdstuk 3 aangevuld met additieve vaste individuele effecten. In tegenstelling tot de interpretatie van het verwijderen van het gemiddelde in de lineaire situatie, is de Pooled NLSDV schatter gedefinieerd als de Pooled NLS schatter voor een auxiliaire regressievergelijking die gedefinieerd is als het verschil tussen de originele regressievergelijking en de vergelijking van het gemiddelde over tijd. Deze auxiliaire vergelijking definieert een andere doelfunctie dan die gedefinieerd door de originele regressievergelijking, maar handhaaft de parameters die van belang zijn. Daarnaast is de auxiliaire regressievergelijking altijd correct door de wiskundige constructie. Daarom is de Pooled NLSDV schatter eigenlijk een Pooled NLS schatter voor een ander regressiemodel. Zo doende heeft de Pooled NLSDV schatter dezelfde convergentiesnelheid als de Pooled NLS schatter als $T \rightarrow \infty$ gevolgd door $n \rightarrow \infty$. Als de niet-lineaire functie H – *regular* is, kan de Pooled NLSDV schatter een sterkere verzameling van identificatiecondities vereisen dan de Pooled NLS schatter. Van de andere kant, als de functie I – *regular* is, heeft de Pooled NLSDV schatter dezelfde asymptotische eigenschappen als de Pooled NLS schatter. Dit bevestigt de conclusie van Chang et al. (2001) dat functies die I – *regular* zijn, asymptotisch orthogonaal zijn ten opzichte van de deterministische componenten. De simulatiestudie gebaseerd op het Logistic Smooth Transition model bevestigt de toepassingen in González et al. (2005) en Fok et al. (2005), als de daar gebruikte data worden beschouwd als $I(1)$.

In Hoofdstuk 5 bekijken we een persistente error multifactor structuur voor panel polynomiale co-summable regressiemodellen. We breiden de CCEP schatter van Pesaran (2006) uit om de parameters van belang te schatten. We onderscheiden twee situaties, te weten homogene en heterogene parameters over i . Voor ieder van deze situaties bekijken we de effecten van zowel de aanwezigheid als afwezigheid van cross-sectionele cointegratie tussen de regressoren. Onze heuristische analyse en de

simulatiestudie geven de indruk dat de aangepaste CCEP schatter nog steeds consistent is met snelheid \sqrt{n} en asymptotisch normaal verdeeld als de coëfficiënten heterogeen zijn over i . Verder wijzen de simulatieresultaten erop dat de convergentiesnelheid van T lijkt af te hangen als de coëfficiënten homogeen zijn.

Curriculum Vitae

Lei Wan was born on May 23, 1977 in Qingdao, China. He obtained a Bachelor of Engineering degree in 2000, and worked for Qingdao Daily Press Group from 2000 to 2003 in China. From 2003 to 2006, he studied Econometrics at Maastricht University, obtaining a Bachelor's degree in Econometrics. During his studies, he spent the fall semester 2005 at the Hong Kong University of Science and Technology, Hong Kong. Since September 2006, he continued his study at Maastricht University, obtaining his M.Phil. in Economic and Financial Research, specialized in Econometrics, with distinction (*cum laude*).

Lei joined the Department of Quantitative Economics at Maastricht University in September 2008. As a Ph.D. student, he worked under the supervision of Prof. dr. Franz C. Palm and Prof. dr. Jean-Pierre Urbain. The results of his research are presented in this thesis.

